January 2010

A Processing Approach to the Working Memory/Long-Term Memory Distinction: Evidence from a Levels-of-Processing Span Task

Nathan Rose
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Examination Committee:
Mitchell Sommers (Co-Chair)
Joel Myerson (Co-Chair)
Joe Barcroft
Pascal Boyer
Henry L. Roediger, III
Denise Wilfley

A PROCESSING APPROACH TO THE WORKING MEMORY/LONG-TERM MEMORY DISTINCTION: EVIDENCE FROM A LEVELS-OF-PROCESSING SPAN TASK

by

Nathan Scott Rose

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2010

Saint Louis, Missouri

ABSTRACT

Recent studies have raised questions about the extent to which working memory (WM) is dissociable from secondary or long-term memory (LTM). Although many similarities may exist between immediate retrieval on WM span tasks and delayed retrieval on LTM tests, important differences exist as well. To illustrate this point, Craik and Tulving's classic levels-of-processing paradigm was adapted for use in a WM span task: Participants made visual, phonological, or semantic judgments about 33 words using the same stimuli and instructions as Craik and Tulving (1975), but were to recall words immediately after every 3 or 8 words (rather than after all words were processed). In the context of this WM span task (Experiment 1), no benefit of deeper processing occurred on immediate recall, even though subsequent recognition of the same items showed the classic levels-of-processing effect. However, when words were processed in the same way but immediate recall was not required (Experiment 2), surprise immediate recall tests did demonstrate a levels-of-processing effect, but only for supraspan (8-item) lists. These results demonstrate both similarities and differences between WM and LTM. One way these disparate effects can be reconciled is within a transfer-appropriate-processing account of the WM/LTM distinction. That is, the WM/LTM distinction depends on the extent to which there is a match (or mismatch) between the processes that are used for initial encoding and subsequent retrieval. For example, when WM tests involved intentional encoding and active maintenance of to-be-remembered words (Experiment 1), a levels-of-processing effect was not observed. However, for surprise recall of supraspan (8-item) lists in Experiment 2, initial processing was not directed at temporary maintenance for immediate recall (because the test came as a surprise), which

made this situation similar to the LTM task. Under these conditions, a levels-of-processing effect (like that observed on LTM tasks) was observed on the WM span task, consistent with a transfer-appropriate-processing account of the WM/LTM distinction.

Acknowledgments

I would like to thank my core dissertation committee members, Joel Myerson, Roddy Roediger, and Mitch Sommers for their guidance and support throughout my doctoral training. I thank Sandra Hale for her guidance and support and many helpful discussions on this dissertation and related work. I thank Gus Craik for many helpful discussions in preparation of this dissertation. I thank members of the Sommers Speech and Hearing Lab, the Hale/Myerson Cognitive Development Lab, and the Roediger Memory Lab for helpful comments as well. I am grateful for financial support from an NIA Institutional Training Grant (PI's: Martha Storandt and David Balota) that funded the present research. I thank Matt Robbins for his assistance with programming the present experiments and collecting, organizing, and analyzing the data. Finally, I thank my wonderful wife, Denise Rose, for her unending love and support, which even included helping me collect some of the data for the present experiments. Your love and devotion have truly surpassed all expectations – thank you.

Table of Contents

Abstract…………………………………………………………………………………ii

Acknowledgments……………………………………………………………………..iv

List of Tables ……………………………………………………………………...vii

List of Figures …………………………………………………………………….viii

Introduction... 1

On the distinction between short-term and long-term memory stores 2

Development of the concept of working memory and its relation to long-term memory 8

Testing whether performance on working memory tasks relies principally on retrieval from secondary memory.. 17

Dissociations between levels-of-processing effects on immediate and delayed memory tests.. 24

Evidence supporting the role of secondary memory on working memory tasks 26

A theoretical interpretation of levels-of-processing effects on working memory and long-term memory 36

The Present Study 40

A processing approach to the working memory/long-term memory distinction 40

Experiment 1: Levels-of-processing effects on immediate recall with intentional encoding and delayed recognition. 41

Methods 43

Results 46

Discussion 52

Experiment 2: Levels-of-processing effects on immediate recall and delayed recognition when testing is unexpected. 57

Methods 59

Results 60

Discussion 63

General Discussion 68

Overview of Findings 69

Implications for the theoretical distinction between working memory and long-term memory 71

Concluding Remarks 79

Appendix A. 80

Appendix B. 85

Appendix C. 86

Appendix D. 87

References 88

List of Tables

Table 1. Examples of typical questions and responses used in Craik and Tulving's (1975) experiments.

Table 2. Mean (SEM) Proportion of Items Correctly Recalled on the Levels-of-Processing Span Task (Intentional Encoding) and Correctly Recognized as Old on the Delayed Recognition Test for Items Initially from 3- or 8-Item Lists.

Table 3. Mean (SEM) Proportion of Items Correctly Recalled on the Levels-of-Processing Span Task Following Incidental Encoding and Correctly Recognized as Old on the Delayed Recognition Test for Items Initially from 3- or 8-Item Lists.

Table 4. Mean proportion of words estimated to have been recalled from primary and secondary memory in Seamon and Murray (1976) and Smith et al. (1971).

List of Figures

Figure 1. Model of memory adapted from Waugh & Norman (1965).

Figure 2 A. Baddeley and Hitch's (1974) model of working memory. **B.** Baddeley's (2000) model of working memory.

Figure 3. Cowan's (1999) embedded process model.

Figure 4. Depiction of Unsworth & Engle's (2007) Dual-Component Model of Working Memory.

Figure 5. The proportion of words recognized as a function of levels of processing in Craik & Tulving, 1975, Experiment 1. "Yes" responses only are displayed.

Figure 6. Procedure of the Levels-of-Processing Span task (Rose et al., 2010). Depicted is an example of an immediate recall test for a 2-item list, the filled delay, and a target word on the surprise recognition test. Note that the immediate recall tests for each subject consisted of all three conditions (color, rhyme, or semantic processing blocked by list) and that each condition consisted of either 2 trials of list lengths ranging from 2-7 items or 3 trials of 4- and 8-item lists.

Figure 7. Immediate recall and delayed recognition results of Rose et al., 2010, Experiment 1.

Figure 8. Delayed recognition data of Rose et al., 2010, Experiment 2: Proportion of words from the LOP span task recognized as 'old' for the group that performed the LOP span task with immediate testing and the group that made the same processing decisions but did not have immediate tests.

Figure 9. Immediate and delayed recall data from Rose et al. (in preparation).

Figure 10. The proportion correct for word from 4- or 8-item lists on immediate recall and delayed recognition (Panel A, from Rose et al., 2010, Experiment 2) or delayed recall (Panel B, from Rose et al. (in preparation), collapsed across level of processing.

Figure 11. Mean proportion of words recalled on the immediate tests and recognized as old target words on the delayed recognition test as a function of level of processing.

Figure 12. Mean proportion of words recalled on the immediate tests for 3- and 8-item lists and mean proportion recognized as old target words on the delayed recognition test for words initially from 3- and 8-item lists.

Figure 13. Immediate recall as a function of level of processing when immediate recall tests were expected (intentional encoding) and unexpected (incidental encoding). Data are estimated from Marsh, Sebrechts, Hicks and Landau (1997) Figure 1B.

Figure 14. Mean proportion of words recalled on the immediate tests for Experiment 1 (immediate tests expected) and Experiment 2 (immediate tests unexpected) as a function of level of processing. Error bars represent the standard error of the mean.

Figure 15. Mean proportion of words recognized (hits) on the subsequent delayed recognition tests when the initial immediate tests were expected (Experiment 1) and when they were unexpected (Experiment 2) as a function of level of processing, collapsed across list length.

Introduction

The idea that short-term and long-term memory represent distinct memory systems has a long history. Recent theories, however, suggest that retrieval from long-term memory is involved in performing short-term memory tasks, including the subset of such tasks that are also known as working memory tasks (e.g., Baddeley, 2009). The present study examined whether performance on working memory and long-term memory tasks are affected similarly by specific manipulations. If the same principles do characterize performance on working memory and long-term memory tasks, then one might expect a levels-of-processing manipulation (Craik & Lockhart, 1972) to affect performance on working memory tasks just as it does performance on long-term memory tasks. In Experiment 1 of the current study, I show that the level of processing during initial encoding does not affect working memory performance. However, there are other pieces of evidence that suggest that retrieval from long-term memory is involved in performance on working memory tasks. Thus, an important theoretical question arises. Why, if retrieval from long-term memory is involved, do working memory tasks fail to show a levels-of-processing effect?

Here I propose that these results may be interpreted within the transfer-appropriate-processing framework of memory (Morris, Bransford, & Franks, 1977). Because working memory tasks involve maintenance of a relatively small set of information over short retention intervals whereas long-term memory tasks typically involve retention of a much larger set of material over longer intervals different processes may be involved in performance on the two types of tasks. That is, even if the same memory system is involved in performance on working memory and long-term memory tasks, the difference in task demands may result in differences in the encoding, maintenance, and/or retrieval processes that are used to perform the two types of tasks, in which case, differences in the effects of many variables (e.g., levels of processing) are to be expected. However, the transfer-appropriate-processing framework also would predict that if the task demands could somehow be made more similar, then the processes involved would be more similar, and some of these differences in the effects of variables should be reduced or even eliminated. Put simply, the transfer-appropriate-processing account of the distinction between working memory and long-term memory would predict that a levels-of-processing effect could be obtained on a working memory task if the processes involved in performance were similar to those involved in performance on long-term memory tests. Experiment 2 of the current investigation tests this hypothesis.

On the distinction between short-term and long-term memory stores

The idea that there is a short-term memory store that is limited in capacity and is different from a long-term store has a long history. Ebbinghaus (1885/1964) reported that he could recall 7 nonsense syllables perfectly after one presentation but that 8, 9, or 10 syllables required more repetitions before they could be perfectly recalled. In 1890, based purely on introspection, William James distinguished between primary and secondary memory. Primary memory was said to reflect the current contents of consciousness, whereas secondary memory was said to consist of memory of the past that must be brought back into consciousness by a retrieval process.

This distinction was maintained in early information-processing models of memory developed by cognitive psychologists (e.g., Atkinson & Shiffrin, 1968;

Broadbent, 1958; Waugh & Norman, 1965). For example, Waugh and Norman's (1965) model (depicted in Figure 1) suggested that when one perceives a stimulus (such as a phone number or someone's name that one has just met) it is quickly forgotten if it is not rehearsed in primary memory. The rapid forgetting of information just perceived is a phenomenological experience to which everyone can attest. Furthermore, Waugh and Norman suggested that information must be sufficiently rehearsed in primary memory in order for the information to be transferred to the more permanent, secondary memory store.

Waugh and Norman's (1965) idea was that a stimulus first enters primary memory (and does not make direct contact with secondary memory) and that one must rehearse information in primary memory in order to transfer the information to secondary memory. Although this idea has received considerable criticism, for present purposes the key point is that many models of memory assume there are two distinct stores: one that is dedicated to maintenance of a rather small set of information over the short term and another that is dedicated to retention and retrieval of information over the long term.[1]

[1] A variety of terms have been used to describe a short-term or temporary memory store as distinguished from a more long-term or permanent store. The terms primary, short-term, working, secondary, and long-term memory have all been used to refer to theoretical constructs in various theories. Perhaps unfortunately, they are also used as adjectives to refer to tasks (e.g., working memory task, long-term memory task, etc.). The problem is that these tasks may rely not only on a single putative memory system. For example, many researchers now believe that both primary and secondary memory are involved in working memory tasks. To minimize confusion, I will use the terms primary and secondary memory to refer to theoretical constructs (except in places where previous authors used other terms) and short-term memory, working memory, and long-term memory to describe different types of tasks.

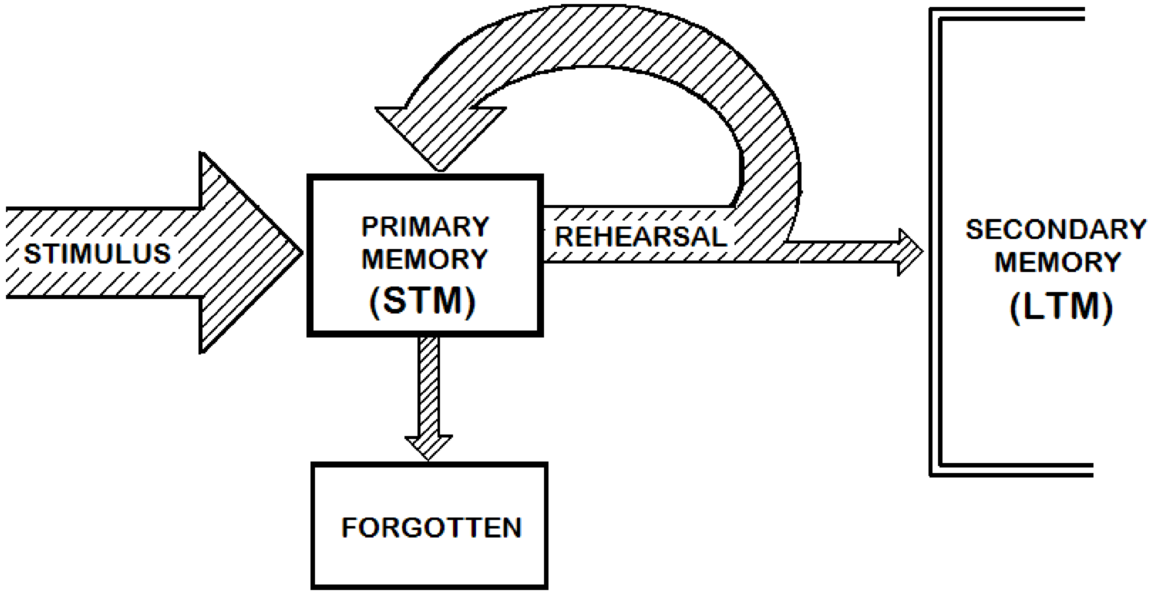


Figure 1. Model of memory (adapted from Waugh & Norman, 1965).

Several lines of evidence support the idea that memory over the short- and long-term involve separate stores. One concerns the difference in the amount of information that can be maintained. Primary memory is limited in capacity in that there is a finite number of items that can be maintained in conscious awareness at any given time. For example, Miller (1956) proposed a "magic number" of 7 plus or minus 2 items, such as digits in a phone number that could be reliably maintained and reproduced over the short-term. Beyond that number, errors are produced which suggests primary memory capacity is exceeded. In contrast, the capacity of secondary memory is assumed to be virtually limitless. Theoretically, researchers assume that humans are capable of storing an endless amount of information over the long term. Whether we can later access that information by retrieving it or not is an entirely different issue (e.g., Tulving & Pearlstone, 1966). Of importance here is the notion of a difference in capacity limitations between short-term and long-term stores. I return to this issue, and the debate surrounding it, below.

Another critical source of evidence for the existence of two separate stores is provided by cases of amnesia following brain damage. Damage to the hippocampus and/or surrounding areas of the medial temporal lobe produces an inability to form or retrieve new long-term memories, as in the famous case of patient H.M. (Milner, 1966). Moreover, patients with amnesia have been reported to have a preserved ability to maintain and reproduce a small subset of information over the short-term. In contrast, patients with damage to perisylvian cortex, such as patient KF, have been reported to show the reverse pattern of impairment: preserved performance on long-term memory tasks, but impaired performance on short-term memory tasks (Shallice & Warrington, 1970). As discussed below, however, more recent studies have raised questions concerning this double dissociation.

Another point of distinction between primary and secondary memory concerns differences in the type of encoding, maintenance and retrieval processes involved in performance on tasks thought to tap the two systems. For example, on short-term or working memory tasks that require remembering a series of words, people tend to rehearse the words, and their performance is better when they can do so without distraction (e.g., the articulatory suppression effect; Baddeley, Lewis, & Vallar, 1984). In contrast, on long-term memory tests, it is usually not possible to rehearse a long list after only a single presentation or to continuously rehearse even a short list over a long delay. Instead, people perform better on long-term memory tests when deeper (semantic) cues are encoded at the time of initial learning than when shallower (perceptual) cues are encoded (i.e., the levels-of-processing effect; Craik & Lockhart, 1972).

Further evidence in support of the distinction between primary and secondary memory comes from serial position effects in list learning experiments (e.g., Baddeley & Warrington, 1970; Craik, 1968; Crowder, 1968). When supraspan lists of to-be-remembered items are recalled, items from the recency portion of the list (i.e., the most recently presented items) are assumed to be reported from primary memory because the items were just perceived. Items from the initial (primacy or pre-recency) portion of the list are assumed to be retrieved from secondary memory because of the distance between the time of encoding and retrieval. Consistent with this view, numerous variables affect retrieval of items from one part of the serial position curve while leaving the other part unaffected.

For example, some variables that benefit recall of pre-recency (i.e., secondary memory) items but not recency (i.e., primary memory) items are deeper levels of processing (Seamon & Murray, 1976; Smith, Barresi, and Gross, 1971), list length (Deese, 1960), the rate at which items are presented (Murdock, 1962; Glanzer and Cunitz, 1966), word frequency (Deese, 1959), semantic similarity (i.e., relatedness, Tulving & Patterson, 1968), and imageability (Paivio et al, 1969). Some variables that negatively affect recall of pre-recency, but not recency, items are participants' age (Craik, 1968), and damage to the medial temporal lobe (Baddeley & Warrington, 1970). There are also many variables that are known to affect the recall of recency (primary memory) but not pre-recency (secondary memory) items. For example, recall of the most recently presented items is negatively affected by a filled delay or retention interval (Brown, 1958; Glanzer & Cunitz, 1966; Peterson & Peterson, 1959; Postman & Phillips, 1965), the modality in which items are presented (visual < auditory; Conrad & Hull, 1964, 1968), the presentation of a stimulus that follows the final item, such as the experimenter saying the word 'recall' (i.e., the suffix effect; Crowder, 1968; Roediger & Crowder, 1976), and lesions to the parietal and temporal lobes (Shallice & Warrington, 1970).

To summarize, performance on short-term and long-term memory tests demonstrate differences in memory capacity, the effects of brain damage, and the processes that lead to better performance. All of these differences are consistent with the hypothesis that there are distinct systems responsible for short-term and long-term storage. Importantly, evidence that manipulations have different effects on immediate recall of items from the recency (primary memory) and pre-recency (secondary memory) portions of a supraspan list strongly suggest that two different "systems" can both be involved in the same task. Below I discuss how a similar hypothesis has been proposed in an attempt to understand performance on working memory tasks.

Development of the concept of working memory and its relation to long-term memory

Historically, the concept of working memory may be thought of as evolving out of the concept of short-term or primary memory. Whereas short-term memory was assumed to be devoted solely to the temporary storage of information, the concept of working memory was developed to capture a more dynamic system in order to explain performance on tasks requiring the simultaneous engagement of processing activities in addition to temporary storage. For example, clearly something more than just temporary storage of information is needed to perform complex cognitive activities such as language comprehension, mathematics, and reasoning. As the concept of working memory developed, however, there was a theoretical shift from the way researchers conceptualized the distinction between working memory and long-term memory. Whereas the distinction between short-term and long-term memory was once quite clear, the division between working memory and long-term memory is considerably less well specified. Below, I discuss this theoretical transition.

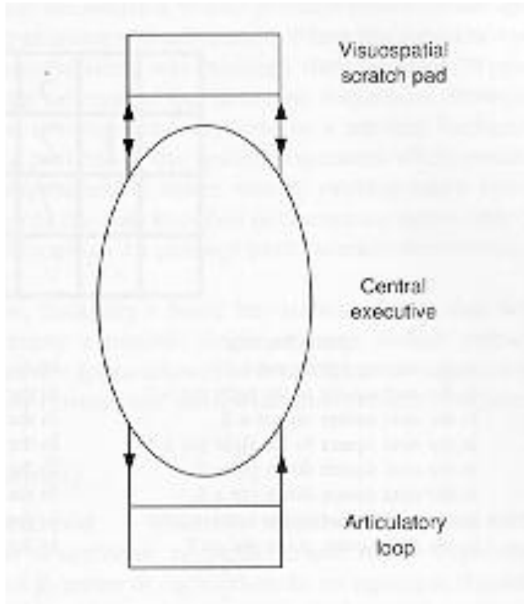
Baddeley and Hitch (1974) proposed the first model of working memory, and Baddeley expanded upon this model in his seminal book (Baddeley, 1986), which included a central executive component responsible for controlling attention during the performance of a task in addition to a set of temporary storage buffers dedicated to the maintenance of particular types of information (verbal information – articulatory loop;

visuospatial information – visuospatial scratch pad). The components of the original model are depicted in Panel A of Figure 2.

Baddeley's (1986; Baddeley & Hitch, 1974) model has dominated the field of working memory research, but his original model did not clearly address some theoretical issues concerning the ways in which working memory and long-term memory are related [issues which Baddeley describes in his recent book (Baddeley, 2009) as "skeletons in the working memory cupboard" p. 114]. Originally, Baddeley's (1986) model maintained a distinct separation between working memory and long-term memory. More recently, he has modified his model (Baddeley, 2000) by adding a component called the "episodic buffer" in acknowledgement of the ways in which working memory and long-term memory interact (see Panel B of Figure 2 for a depiction of Baddeley's revised model).

Baddeley (2009) noted that there are many ways in which working memory and long-term memory interact. For example, maintaining information that one is already familiar with (e.g., words or symbols which already have representations in long-term memory) is easier than maintaining novel information (e.g., nonwords or novel shapes). Furthermore, chunking bits of information together that one is already familiar with is known to benefit performance on both short-term and long-term memory tasks (Miller, 1956). For instance, a short-term memory task might require one to remember a series of letters for immediate serial recall, such as "*n – i – m – h – n – i – h – n – s – f.*"

A



B

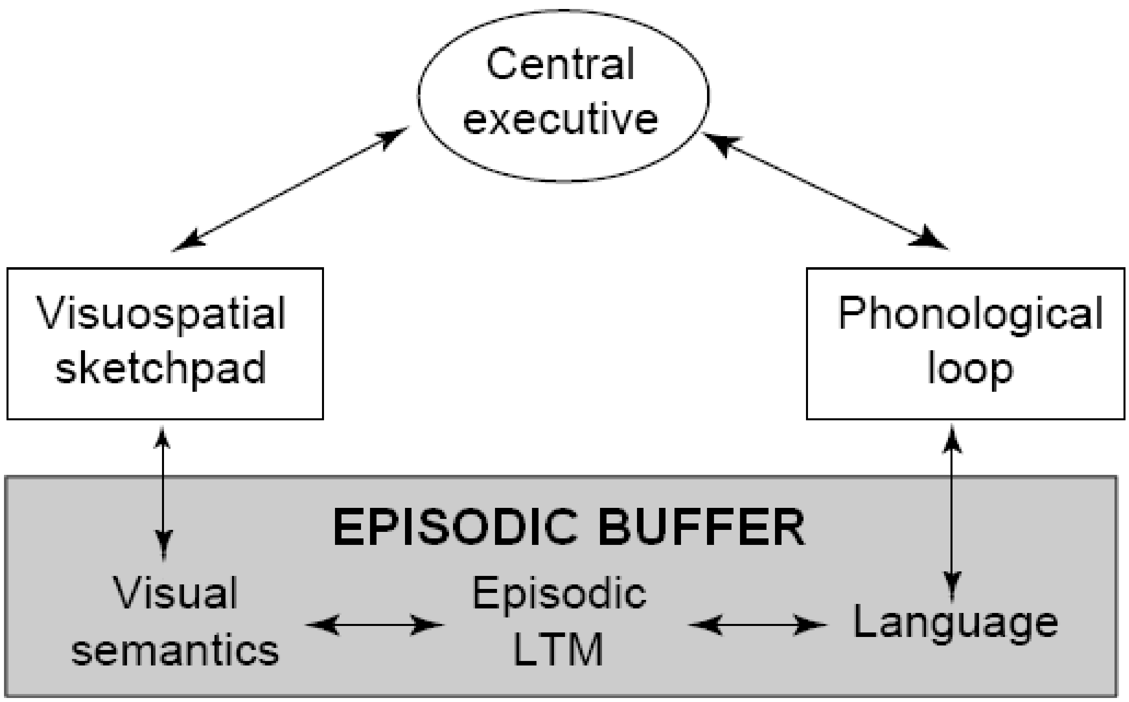


Figure 2 A. Baddeley and Hitch's (1974) model of working memory. **B.** Baddeley's (2000) model of working memory.

Although such a long series of letters exceeds short-term memory capacity for most individuals, the task becomes considerably easier for many American psychologists if they “chunk” the items together into meaningful units, as in *NIMH, NIH, NSF*. However, such chunks would not necessarily help those unfamiliar with American research funding agencies (e.g., the National Institute of Mental Health, the National Institutes of Health, and the National Science Foundation) and who do not have these acronyms already stored in long-term memory.

In Baddeley’s (2000) revised model, he proposes that the episodic buffer component is responsible for “binding” items together into meaningful units and maintaining cues that can be used to access these meaningful representations stored in long-term memory. When people chunk bits of information together by establishing meaningful associations between individual items, the items are grouped into a bound representation that is already stored in long-term memory (e.g., the association between the individual letters N, I, M, and H and the meaningful chunk NIMH). The episodic buffer maintains cues (e.g., “a funding agency”) for these associations. Thus, Baddeley now believes that an important function of his proposed working memory system is to act as an “interface” between the primary and secondary memory systems. This, of course, begs the question of whether the construct of a working memory system is needed, or whether performance on working memory tasks may be more simply described in terms of the interaction of the primary and secondary memory systems, without the need to invoke a distinct working memory system at all.

There are several other models of working memory that also posit distinct short-term and long-term memory stores involved in performance of working memory tasks, although these models vary greatly in terms of how the short-term and long-term memory stores are assumed to interact. For example, Cowan's (1999) "embedded process model" proposes that information in long-term (secondary) memory may exist in varying states of accessibility based on their level of activation. Items that have been recently perceived or processed, or that are currently being attended to, are activated representations. According to Cowan's model of the human memory system, only a small number (i.e., four) of such activated representations or "chunks" can be held within a component that is aptly termed the "focus of attention," and which is virtually synonymous with primary memory. So long as these items are in the focus of attention, their level of activation does not decay. In contrast, the level of activation of items outside the focus of attention is subject to decay. However, attention can be refocused on these items to restore their level of activation and accessibility. Thus, rather than being structurally distinct, Cowan views working memory as a subset of long-term memory (see Figure 3 for a depiction of Cowan's embedded process model).

Oberauer's (2002) model of working memory is similar to Cowan's in that memory items may exist in varying states of accessibility. Recently processed items have the most activated representations and are immediately accessible. However, Oberauer disagrees with Cowan in terms of the capacity of this component. According to Oberauer, only one item or chunk may be focused on at any given time – not four. Nevertheless, Oberauer does acknowledge that recently activated items or chunks may have privileged accessibility relative to items in the inactive portion of long-term memory. Most importantly for current purposes, both researchers agree that items maintained in working memory are an activated subset of long-term (secondary) memory.

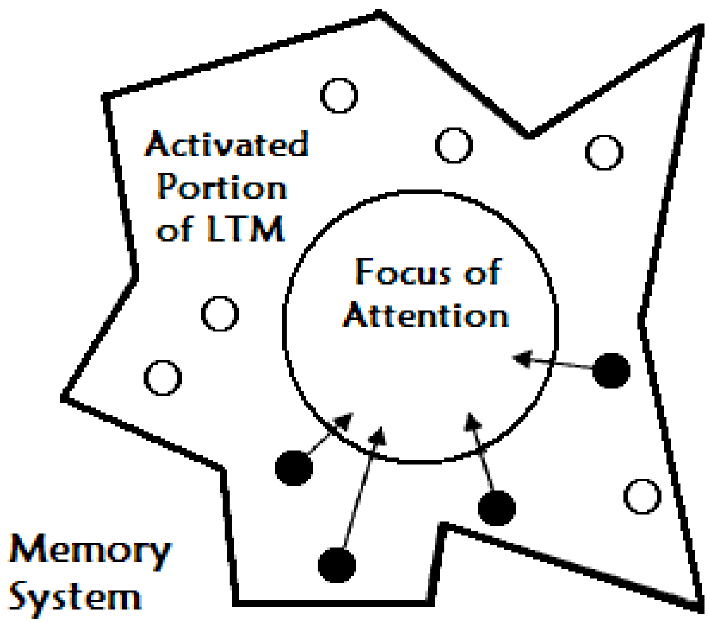


Figure 3. Cowan's (1999) embedded process model. LTM = Long-Term Memory.

Similarly, Unsworth and Engle's (2007) recent dual-component model of working memory also suggests a large amount of overlap between the nature of performance on working memory and long-term memory tasks. Notably, Unsworth and Engle have reintroduced the concepts of primary and secondary memory to the debate by proposing that working memory relies on both systems. That is, a small number of items (e.g., 1 to 4) may be simultaneously maintained within primary memory (or the focus of attention, e.g., Cowan, 1999). When primary memory capacity has been exceeded, Unsworth and Engle suggest that retrieval from secondary memory is required, even though the time between encoding and retrieval is not as long as in traditional long-term memory tasks.

Unsworth and Engle have proposed that different types of immediate recall tasks engage primary and secondary memory to varying degrees. For example, simple span tasks (sometimes called short-term memory tasks), such as digit span, capture the ability to maintain a list of items (e.g., digits) and report them directly from primary memory. This is the case unless the list exceeds approximately four chunks, at which point both primary and secondary memory components are involved (see Unsworth & Engle, 2006). In contrast, complex span tasks (sometimes called working memory span tasks), such as reading span or operation span, require participants to perform a secondary processing task (e.g., reading sentences, solving math problems, etc.) interleaved between presentation of to-be-remembered items. According to Unsworth and Engle's dual-component model, such secondary tasks require that participants temporarily switch attention away from maintaining the to-be-remembered items in primary memory. Therefore, although a few items may be reported from primary memory, at least some of the items must be retrieved from secondary memory, even though recall is relatively immediate compared to most long-term memory tasks (Unsworth & Engle, 2007).

According to Unsworth and Engle, although both primary and secondary memory are involved in performing both simple and complex span tasks, the simple span tasks rely much more on primary memory (for lists of approximately 4 items or less) while complex span tasks rely for the most part on secondary memory. Figure 4 presents Unsworth and Engle's (2007) dual-component model of working memory, depicted so as to facilitate comparison with the Waugh and Norman (1965) model (Figure 1).

In sum, theorists currently differ in how they conceptualize working memory and, in particular, how it is to be distinguished from long-term or secondary memory. Early models (e.g., Waugh & Norman, 1965) made clear distinctions between short-term and long-term memory stores. More recently, however, there is growing consensus that, because the capacity of primary memory is so limited, working memory tasks mostly involve retrieving information from long-term (secondary) memory (Baddeley, 2000; Cowan, 1999; Oberauer, 2000; Unsworth & Engle, 2007).

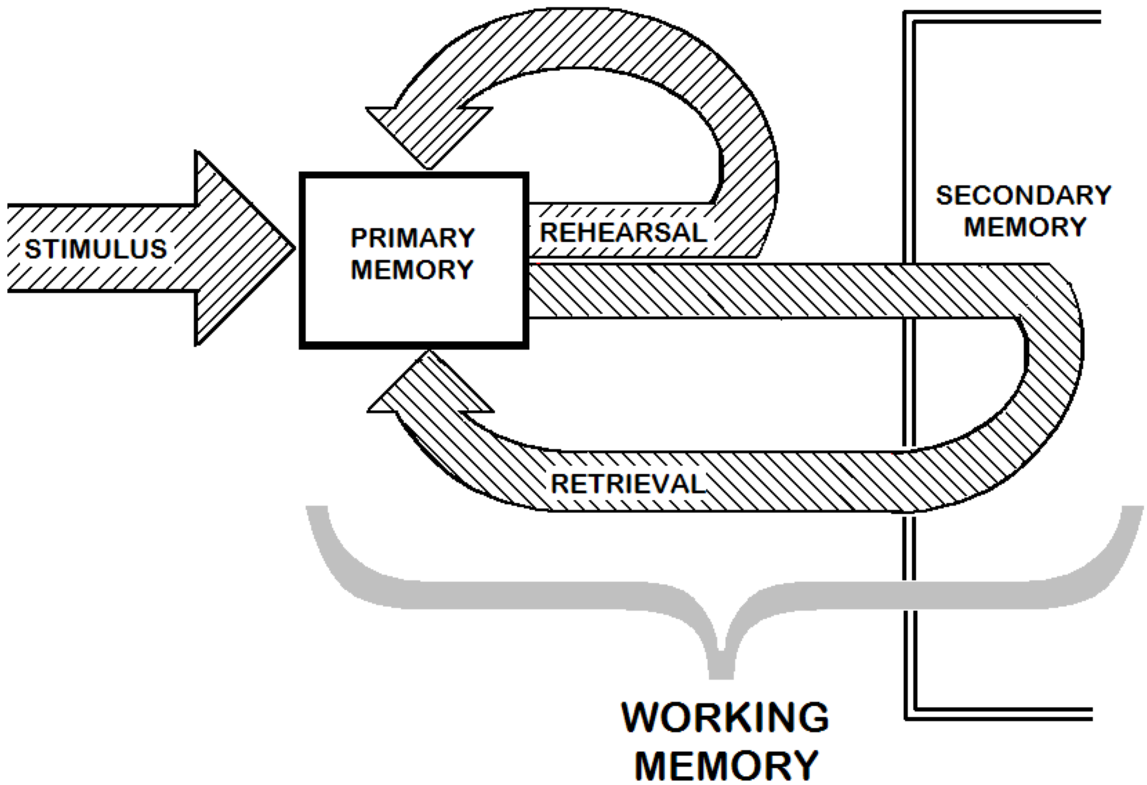


Figure 4. Depiction of Unsworth and Engle's (2007) Dual-Component Model of Working Memory.

Testing whether performance on working memory tasks relies principally on retrieval from secondary memory

The idea that performance on working memory tasks principally involves retrieving items from long-term memory represents a stark departure from previous conceptualizations of the short-term/long-term memory distinction. For example, estimates about the size of primary memory capacity have systematically shrunk since Miller's (1956) early proposal of 7 ± 2 items. Cowan (1999) proposed the number to be 4 (although so too did Watkins, 1974). Oberauer (2002), McElree (2001), and others have proposed the number is actually closer to 1. It should be noted that if only 1-4 items can be reported directly from primary memory, then a task in which one has to recall items from an 8-item list, as in the current study, would rely for the most part on retrieval from secondary memory. Thus, as the hypothesized capacity of primary memory has shrunk, the theoretical importance of retrieval from secondary memory for performance on working memory tasks has increased. Consistent with this emerging view of the role played by secondary memory on working memory tasks, recent neuropsychological case studies of people with amnesia and functional neuroimaging studies of healthy adults have shown that working memory tasks, like long-term memory tasks, depend upon the hippocampus and the medial temporal lobe (see Jonides, Lewis, Nee, Lustig, Berman, & Moore, 2008, for a review), regions that are usually associated with long-term or secondary memory.

If performance on working memory tasks is largely tapping secondary memory, then it would seem to follow that experimental manipulations should produce patterns of effects on working memory tasks similar to those observed on long-term memory tasks. However, the results of a recent study (Rose, Myerson, Roediger, & Hale, 2010) showed that attending to different types of features (e.g., visual, acoustic, semantic) of words at encoding produced effects on working memory tasks that were *not* similar to the pattern (i.e., levels-of-processing effects) that is typically observed on long-term memory tasks. Below, I describe the levels-of-processing effect and review research showing dissociations between levels-of-processing effects on immediate (e.g., working memory) and delayed (long-term memory) tests.

The levels-of-processing framework proposed by Craik and Lockhart (1972) and initially tested in a series of ten experiments by Craik and Tulving (1975) suggests that performance on long-term memory tasks is highly sensitive to the qualitative level or "depth" with which memory items are processed when they are initially encoded. Semantic or conceptual processing at encoding was found to produce superior long-term retention, relative to processing that focused on more structural or perceptual aspects of the memory items, such as phonological or visual features. For example, Craik and Tulving (1975, Experiment 1) presented a series of questions that oriented the processing of individual words (as depicted in Table 1). Following this encoding phase, delayed long-term memory tests (free recall and recognition) showed a substantial benefit of deeper levels of processing (see Figure 5).

Table 1. Examples of typical questions and responses used in Craik and Tulving's (1975) experiments.

Level of processing	Question	Response	
		Yes	No
Structural	Is the word in capital letters?	TABLE	Table
Phonemic	Does the word rhyme with: WEIGHT?	Crate	MARKET
Category	Is the word a type of fish?	SHARK	Heaven
Sentence	Would the word fit in the sentence: He met a ..... in the street?	FRIEND	Cloud

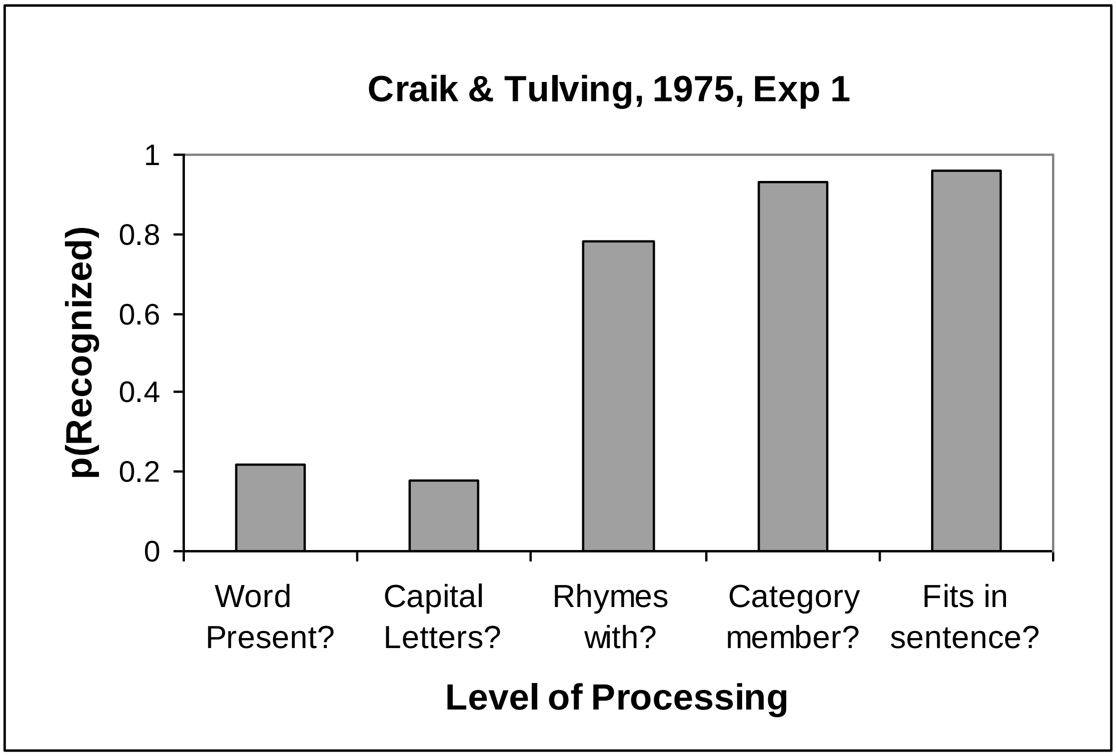


Figure 5. The proportion of words recognized as a function of levels of processing in Craik & Tulving, 1975, Experiment 1. “Yes” responses only are displayed.

Rose et al. (2010) tested whether the type of processing at encoding would affect performance on a working memory span task in the same way that it affects performance on long-term memory tests. In order to conduct this test, Rose et al. developed a levels-of-processing (LOP) span task in which, in addition to the primary immediate recall task, participants performed a secondary task that involved making semantic, phonological, or visual judgments. These judgments involved matching to-be-remembered target words with a semantic associate, a rhyme, or a word presented in the same color, thereby orienting participants as to the level at which they should process the target words (see Figure 6 for a depiction of the procedure). Following the portion of the experiment in which participants performed a series of immediate serial recall tests as part of the LOP span task, they then solved math problems for 5 or 10 minutes, after which they took a delayed recognition test involving all of the target words from the LOP span task and an equal number of new words.

Rose et al. (2010) reported the results of three experiments, in none of which did the level of processing at encoding affect immediate recall on the working memory task (i.e., the LOP span task; see Immediate data in Figure 7). In contrast, long-term memory (assessed by recognition of the same words after a 5-10 minute delay) demonstrated the classic levels-of-processing effect (see Delayed data in Figure 7).

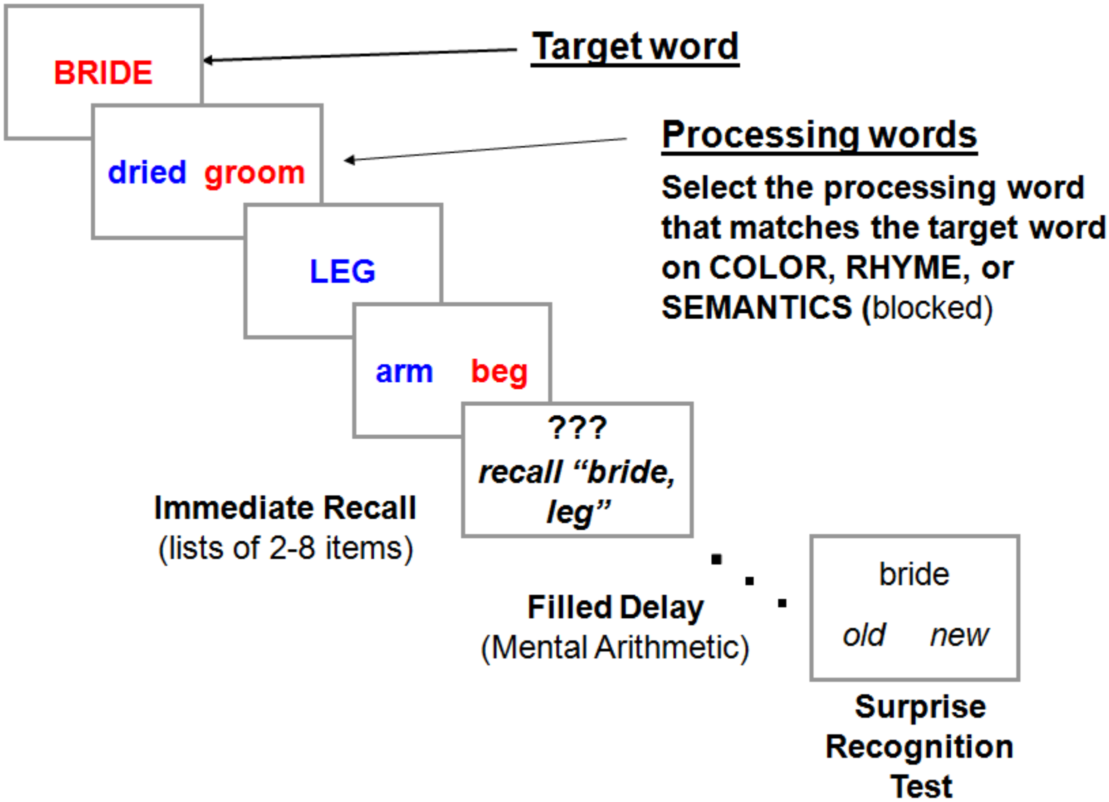


Figure 6. Procedure of the Levels-of-Processing Span task (Rose et al., 2010). Depicted is an example of an immediate recall test for a 2-item list, the filled delay, and a target word on the surprise recognition test. Note that the immediate recall tests for each subject consisted of all three conditions (color, rhyme, or semantic processing blocked by list) and that each condition consisted of either 2 trials of list lengths ranging from 2-7 items or 3 trials of 4- and 8-item lists.

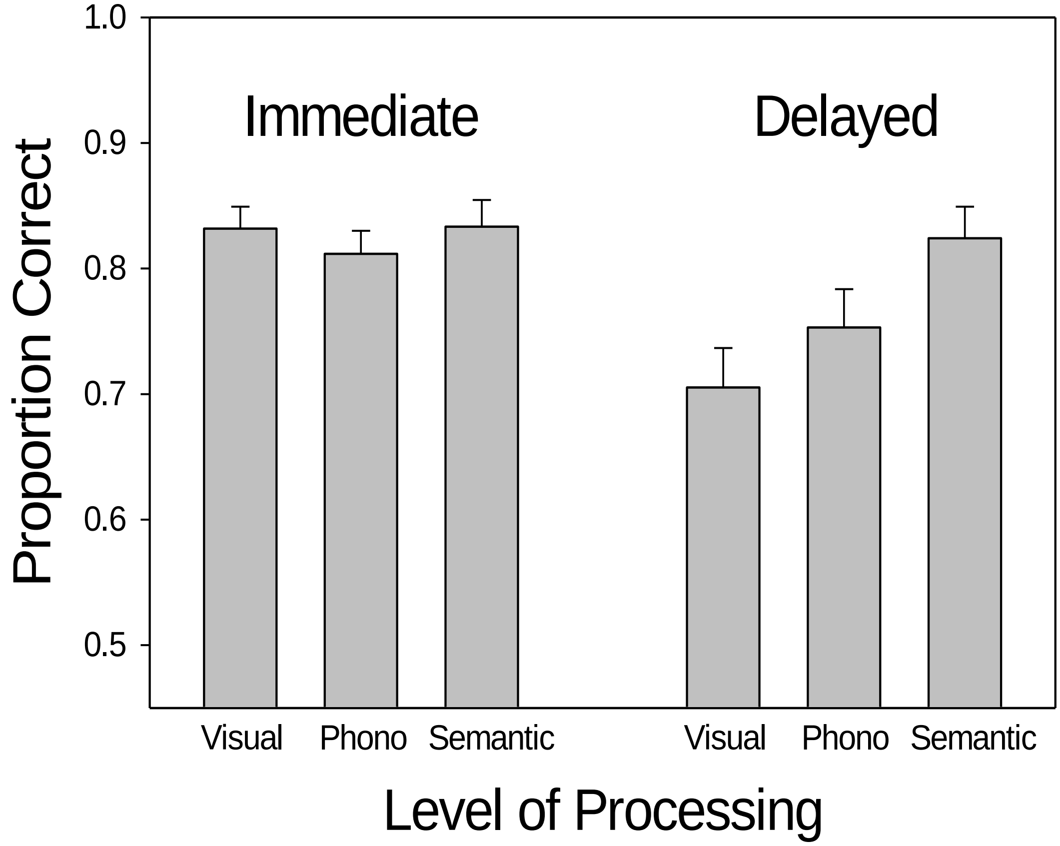


Figure 7. Immediate recall and delayed recognition results of Rose et al., 2010, Experiment 1.

These results clearly show that the same levels-of-processing manipulation that affected long-term retention failed to affect performance on the LOP span task (Rose et al., 2010). One possibility, however, is that these findings were somehow due to the fact that working memory performance was assessed using recall, whereas recognition tests were used to measure long-term memory. To address this potential methodological confound, a follow-up study was conducted in which participants performed the phonological and semantic processing conditions of the LOP span task, as well as a simple word span task, followed by delayed free recall of all the words from the initial span tasks. Similar to our previous findings, there was no difference in immediate recall between the phonological and semantic processing conditions of the LOP span task, but delayed recall of the same items was enhanced for words that were semantically processed on the initial working memory tests (Rose, Myerson, & Roediger, in preparation). These results show that the dissociation between levels-of-processing effects on working memory and long-term memory tests observed previously (Rose et al., 2010) were not due to the use of recognition procedures for assessing long-term memory.

Dissociations between levels-of-processing effects on immediate and delayed memory tests.

The finding that the level of processing at encoding affects long-term retention but does not affect immediate recall on a working memory task would seem to be inconsistent with the idea that retrieval from secondary memory is involved in performance on working memory tasks. One might think that if performance on working memory tasks is largely tapping secondary memory, then experimental manipulations should produce similar effects on working memory and long-term memory tasks.

In this section I review studies that have shown dissociations between levels-of-processing effects on a variety of immediate tests (that are quite different from working memory procedures) and a variety of long-term memory tests. These studies point to the robustness of levels-of-processing dissociations between immediate and delayed memory tests.

A particularly telling example comes from a study by Mazuryk and Lockhart (1974) in which participants were presented with series of five words for immediate free recall. They were instructed that, following presentation of each to-be-remembered word, they were to process that word in one of four different ways, depending on the condition: Either rehearse the word silently, rehearse the word overtly, generate a rhyme (shallow processing), or generate a semantic associate (deep processing). The two rehearsal conditions both produced near perfect immediate recall, but significantly poorer performance was observed in the two conditions with a secondary processing demand (rhyme or semantic generation). Interestingly, the latter two conditions, which most closely resemble the LOP span task with deep versus shallow processing requirements, failed to show a levels-of-processing effect: Generating a semantic associate (semantic processing) did not produce significantly better immediate recall than generating a rhyme (phonological processing).

Subsequently, participants were given either a delayed free recall test or a delayed recognition test on all of the studied words. Semantic processing, despite producing immediate recall performance that was equivalent to phonological processing and worse than either covert or overt rehearsal, resulted in performance superior to all other conditions on both delayed recall and delayed recognition tests. Taken together, the results of the Mazuryk and Lockhart (1974), although undertaken in a quite different theoretical context, are completely consistent with those of Rose et al. (2010) and Rose et al. (in preparation). As in these two more recent studies, Mazuryk and Lockhart found that, compared to rehearsal, semantic processing did not benefit immediate recall yet enhanced long-term retention, as assessed by both recognition and recall tests.

Moreover, several other studies that have compared depth of processing effects on immediate and delayed tests have also reported results showing that semantic processing did not benefit immediate recall or recognition yet enhanced long-term retention for the same words (Craik, 1973; Jacoby, 1974; Mazuryk, 1974). Thus, the combination of a lack of a levels-of-processing effect on immediate memory tests and the evidence supporting levels-of-processing effects on delayed memory tests of the same items appears to be a reliable finding. The findings of Rose et al. (2010) and Rose et al. (in preparation) show that this pattern is also observed in the context of immediate recall on working memory tasks, even when recall involves lists of words above span (i.e., 8-item lists).

Evidence supporting the role of secondary memory on working memory tasks

The difference in levels-of-processing effects between immediate and delayed memory tests represents a clear dissociation that may seem contrary to the idea that working memory tasks involve retrieval from secondary memory. However, other pieces of evidence, reviewed below, are consistent with the hypothesis that retrieval from secondary memory is involved in performance on the LOP span task. This evidence comes from demonstrations of retrieval practice effects.

Recalling items that are assumed to be retrieved from secondary memory is known to benefit the ability to recall that information on later memory tests. This finding is termed the testing or retrieval practice effect. For example, Roediger and Karpicke (2006) reviewed studies that examined memory performance following conditions in which subjects learned information (e.g., prose passages, word pairs, foreign vocabulary) either through repeated studying or repeated study and test trials. They found that learning information with repeated tests (i.e., retrieval practice) resulted in much greater long-term retention on delayed tests than did repeated studying.

However, retrieval practice does not always benefit subsequent memory. Recalling items that are assumed to be reported directly from primary memory is thought to have little or no benefit for the ability to recall that information on later tests. For example, Madigan and L. McCabe (1971) presented participants with 5 word pairs and administered cued recall tests on one of the pairs immediately following presentation of the fifth pair. Then, after a series of immediate recall tests, they administered a final cued recall test on all of the pairs. They found that, unsurprisingly, immediate cued recall of the fifth pair was perfect as the most recently presented item could be reported directly from primary memory. However, final cued recall of the fifth pair was almost always forgotten. In contrast, initial tests on word pairs from earlier positions enhanced delayed cued recall relative to pairs that were not initially tested. The following quote from Karpicke and Roediger (2007) nicely summarizes the key point: “The critical factor for increasing long-term retention is providing an initial test in which recall is possible but relatively difficult. ... If retrieval occurs from primary memory, there will probably be little advantage in the long term. Maintenance rehearsal is a form of repeated retrieval from short-term memory and provides little or no benefit to recall ...” (p. 706). The point is that retrieval practice does not *uniformly* benefit delayed memory. Rather, the amount of benefit observed depends upon the extent to which items are retrieved from secondary memory.

Thus, if recalling items for the LOP span task involves retrieval from secondary memory, there should be a benefit to long-term memory relative to a condition without initial recall tests. If this were true, it would provide support for the hypothesis that retrieval from secondary memory was involved in performance of the LOP span task.

Rose et al. (2010) provided such evidence that performance of the LOP span task involves retrieval from secondary memory and not just recall from primary memory. First, they tested whether retrieval practice benefited long-term retention, relative to a condition in which subjects made the same levels-of-processing (vowel, rhyme, or semantic) decisions on the words, but did not have to recall the items on immediate memory tests (i.e., no retrieval practice). In this condition, participants were not expecting to have to recall the words because there were no immediate tests. Rather, participants were instructed to simply make the levels-of-processing decisions as fast and as accurately as possible. Then, after processing all of the words and performing a 10 minute distractor task, participants were given a surprise recognition test on the words. This recognition test was the same as that for the group that performed the LOP span task, which did involve immediate testing.

Relative to the group who performed the same processing operations but without immediate testing (i.e., no retrieval practice), performing the LOP span task with immediate testing resulted in better delayed recognition of the words (see Figure 8). As discussed previously, if the words were simply reported from primary memory on the initial working memory tests, then there would have been little or no benefit to the long-term retention of the words, relative to the condition without testing (Roediger & Karpicke, 2006). As can be seen in Figure 8, that was clearly not the case. Having to recall the items for the LOP span task facilitated their long-term retention.

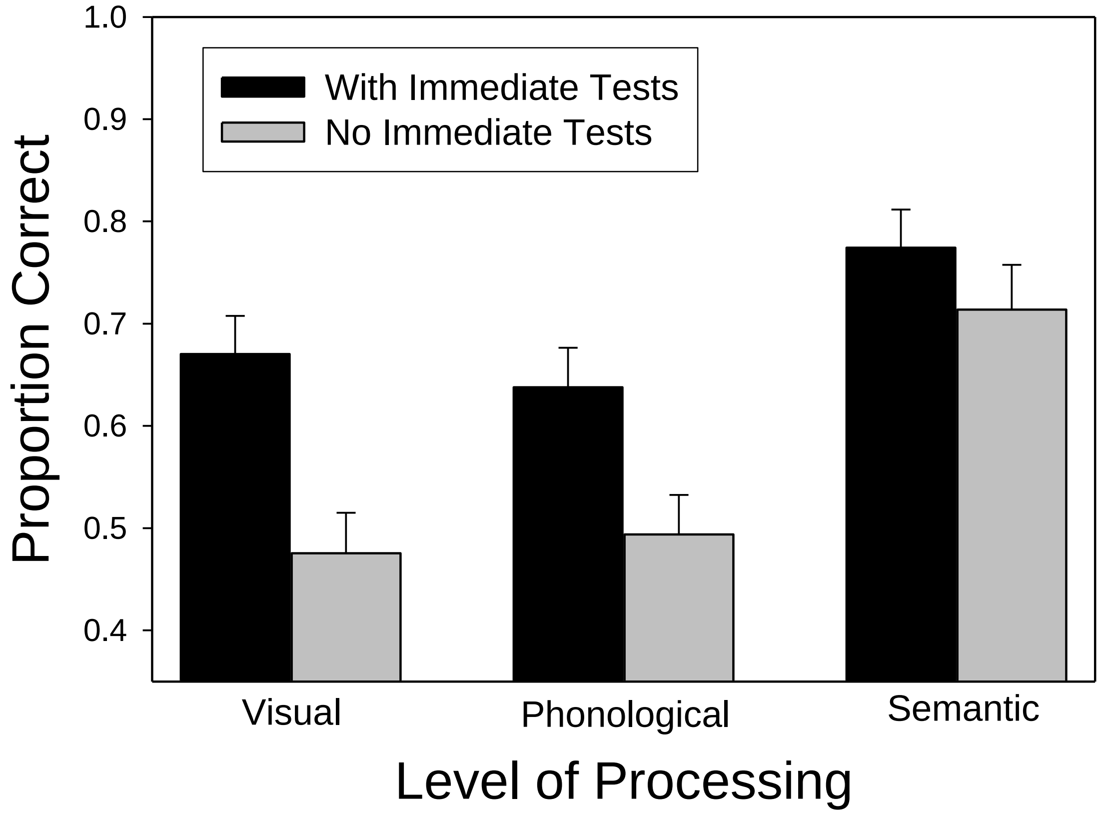


Figure 8. Delayed recognition data of Rose et al., 2010, Experiment 2: Proportion of words from the LOP span task recognized as old for the group that performed the LOP span task with immediate testing and the group that made the same processing decisions but did not have immediate tests.

Therefore, despite the lack of a levels-of-processing effect on immediate recall which would seem to suggest that retrieval from secondary memory was not involved in WM, the retrieval practice effects suggest that retrieval from secondary memory was indeed involved in the immediate recall tests. Recently, we (Rose, Myerson, & Roediger, in preparation) conducted a study to further examine levels-of-processing effects on working memory and long-term memory and the role of retrieval from secondary memory in performance on working memory tasks. We had participants perform both a word span task and the phonological and semantic conditions of the LOP span task. The word span task is a simple span task in which most of the items may be maintained in and reported directly from primary memory (at least for shorter lists). In contrast, the LOP span task involves secondary processing operations and, therefore, should involve retrieving items from secondary memory. If the LOP span task were to involve retrieval from secondary memory, then long-term retention should be better for words recalled on the LOP span task than the word span task.

Immediately following presentation of 4- and 8-word lists, participants had to recall these lists, and this was followed by a final free recall test for all of the words from those span tasks. On the immediate tests participants recalled more items from the word span task than from the complex span task, but the delayed test produced the opposite pattern: Participants recalled *more* items from the LOP span task even though they were *less* likely to recall these items on immediate tests (see Figure 9).

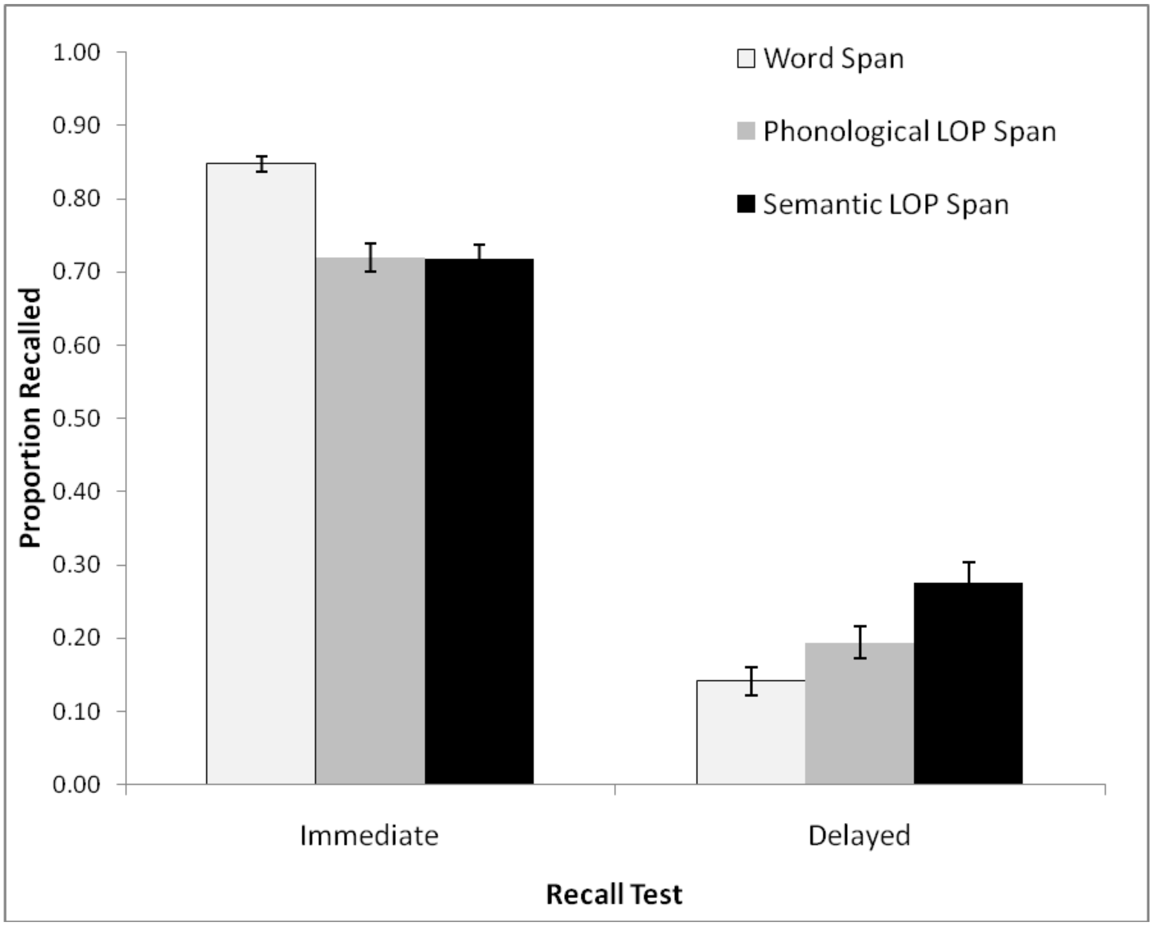


Figure 9. Immediate and delayed recall data from Rose et al. (in preparation).

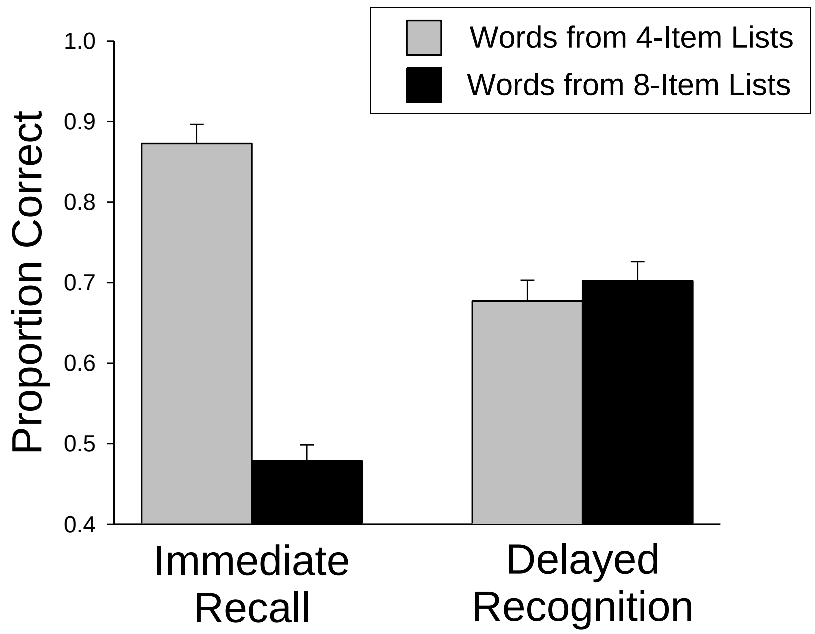
The difference in long-term retention of words initially recalled from the word span and LOP span task may be understood in terms of a difference in the retrieval processes involved in working memory and long-term memory tests. Because the word span task did not involve a distracting secondary task, to-be-remembered items were less likely to have been displaced from primary memory. For the LOP span task, in contrast, following each presentation of a to-be-remembered word, participants were required to process a rhyme and semantic associate of the to-be-remembered word that were not to be recalled. Having to process these not-to-be-remembered words may have displaced the to-be-remembered words from primary memory, and therefore, the words on the LOP span task were more likely to have been retrieved from secondary memory than words from the word span task on immediate recall tests (e.g., D. P. McCabe, 2008; Unsworth & Engle, 2007).

Given that practice retrieving items from secondary memory is beneficial for long-term retention whereas reporting items directly from primary memory has no such benefits (e.g., Roediger & Karpicke, 2006), the differential benefits of retrieval practice for words from the word span and LOP span task likely reflects greater involvement of retrieval from secondary memory on the LOP span task. Immediate recall was best for the word span task but final free recall was poorest for items from this task. Thus, the enhanced long-term retention of items recalled on the LOP span task relative to the word span task strongly suggests that the LOP span task does involve retrieval from secondary memory, despite the absence of a levels-of-processing effect on immediate recall.

Another way in which Rose et al. (2010) and Rose et al. (in preparation) provided evidence in support of the hypothesis that retrieval from secondary memory was involved in performance of the LOP span task was by comparing the long-term retention of items retrieved from short versus long lists. We predicted that retrieval practice would not be as beneficial for items recalled from short (e.g., 4-word) lists as it would be for longer (e.g., 8-word) lists because words from longer lists are more likely to be retrieved from secondary memory than words from short lists. As discussed above, short lists of items may be maintained within and reported directly from primary memory at the time of test (Unsworth & Engle, 2006), and such retrieval is found to have little or no benefit to long-term retention (Roediger & Karpicke, 2006).

Consistent with this hypothesis, words recalled from longer (8-item) lists on immediate tests were less likely to be forgotten later than words recalled from shorter (4-item) lists (see Figure 10). If immediate recall of 4-items lists and 8-items lists were similar, why would the rate of forgetting differ for items initially recalled from 4- and 8-item lists? The difference in rates of forgetting for sub- and supra-span list items provides further evidence that retrieval from secondary memory was involved in performance on the LOP span task, especially for longer lists.

Panel A.



Panel B.

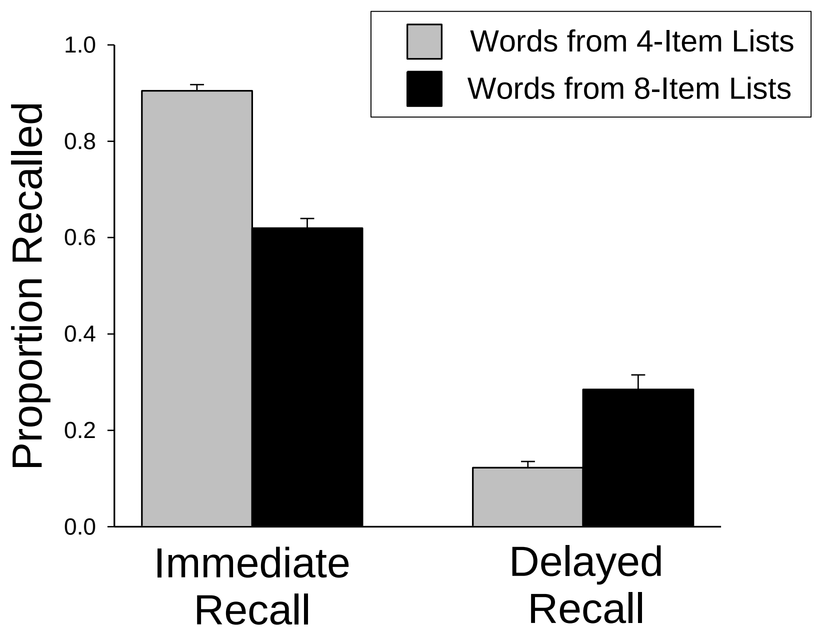


Figure 10. The proportion correct for word from 4- or 8-item lists on immediate recall and delayed recognition (Panel A, from Rose et al., 2010, Experiment 2) or delayed recall (Panel B, from Rose et al. (in preparation), collapsed across level of processing.

In sum, the results of Rose et al. (2010) and Rose et al. (in preparation) demonstrate a stark difference between performance on working memory and long-term memory tasks, in that the two types of tasks are differentially affected by levels of processing. However, other pieces of evidence strongly support the hypothesis that retrieval from secondary memory is involved in performance of working memory span tasks: Retrieving items on working memory tasks benefited long-term retention and it did so to a greater extent when retrieval emphasized secondary memory.

One important theoretical question that remains, however, is if retrieval from secondary memory was involved on the LOP span task, as indicated by the effects of retrieval practice just described, why was there a lack of a levels-of-processing effect on immediate recall? In the following section I consider a potential explanation for these findings.

A theoretical interpretation of levels-of-processing effects on working memory and long-term memory

The finding of a lack of levels-of-processing effects on working memory despite the hypothesized involvement of retrieval from secondary memory (e.g., Rose et al., 2010) may be explained in terms of the transfer-appropriate-processing framework (Morris, Bransford, & Franks, 1977) and the related concept of encoding specificity (Tulving & Thomson, 1973). More specifically, the principles of transfer-appropriate processing and encoding specificity may be applied to the distinction between working memory and long-term memory in the following way: Differences in the pattern of performance between working memory and long-term memory tasks depend on whether different processes – not systems – are involved. If the processes involved in initial encoding and subsequent retrieval were similar between working memory and long-term memory tasks, then a similar pattern of performance should be observed on the two types of tasks.

According to the principle of transfer-appropriate processing, the long-term retention of items is determined not only by the depth of processing at encoding, but by how well the requirements of a subsequent memory test match the processes originally used to encode information. For example, Morris et al. (1977) found that rhyme processing at encoding produced better long-term memory than semantic processing – a reversal of the standard levels-of-processing effect – if the subsequent memory test involved recognizing words that rhymed with the studied words. Similarly, Stein (1978) found that visual processing of words presented in various upper- and lower-case patterns resulted in better long-term memory than semantic processing on a case-oriented recognition test. The memory test required participants to discriminate target words from foils that differed in their case pattern ("when you saw the word *wind*, was it *Wind, wInd, wiNd,* or *winD*?"). According to the transfer-appropriate-processing framework, long-term memory depends on the match between initial encoding and subsequent retrieval. Applied to the current study, it is possible that the lack of levels-of-processing effect on the LOP span task may be due to a mismatch in the processes used between the initial working memory tests and the subsequent long-term memory test.

The processes involved in working memory and long-term memory tests likely differ because the two types of tests have different requirements. Working memory tests require maintaining a small amount of information over the short-term whereas long-term memory tests require maintaining a large amount of information over the long-term. Therefore, the processes involved in working memory tests will tend to be those that are required for maintenance rehearsal and recall over short retention intervals. In contrast, long-term memory tests tend to encourage the use of cue-dependent retrieval processes to recover information about previously encoded episodes. For example, if someone is trying to remember a phone number such as 215-1904 until it can be dialed into a telephone, he or she may try to rehearse the numbers repeatedly if the call will be made within a minute or so. If, however, the call is to be made at a much later time, they may try to establish a more meaningful retrieval cue such as 215 – February 15, a friend's birthday – and 1904 – the year the world's fair was held in St. Louis. These examples illustrate how differences in the requirements of the task (maintain digits to dial into the phone vs. retrieve digits after a long delay) result in differences in the cognitive processes that are involved, even when the to-be-remembered information is the same.

Even if both working memory and long-term memory tasks involve retrieval of the same information from the same secondary memory system, the demands of the two types of tasks will likely bias the use of different processes. With respect to the LOP span task, the secondary tasks not only require that participants process the words in ways that are more or less beneficial for retrieving the words from secondary memory, they also require attending to words that are not to be remembered. Therefore, it is likely that at least some of the to-be-remembered words must be recalled from secondary memory, particularly for supraspan list lengths. Nevertheless, because the LOP span task is a working memory task and thus requires immediate recall, participants also likely engage in maintenance rehearsal to try to maintain the words.

For immediate recall, if words are being actively maintained up until the recall test, then the level of processing at encoding may not be important. Rather, the involvement of maintenance rehearsal processes may be sufficient for performance on working memory tests like the LOP span task. This is particularly true if the task involves maintenance and retrieval of a rather small set of information such as a short list of words, and if storage, retention, and retrieval of a large set of information for a delayed test is not anticipated. This interpretation of the working memory/long-term memory distinction as applied to the Rose et al. (2010) study is consistent with the transfer-appropriate-processing framework (Morris et al., 1977).

From the perspective of the transfer-appropriate-processing framework, differences between levels-of-processing effects on working memory and long-term memory reflect differences in the processes that are involved. If one task shows a levels-of-processing effect and the other does not, then the processes likely differ. This need not be seen as evidence contrary to the idea that performance on working memory tasks involves retrieval from long-term memory. Rather, I hypothesize that if similar processes were involved, levels-of-processing would affect working memory and long-term memory tasks similarly. The present experiments were designed to test this hypothesis.

The goal is for the transfer-appropriate-processing framework to provide a unified account of how working memory and long-term memory may demonstrate both differences (Experiment 1) and similarities (Experiment 2) in the effects of certain manipulations (e.g., levels of processing), despite the involvement of retrieval from long-term memory in both types of tasks. This processing based approach to the working memory/long-term memory distinction may also serve to reconcile the evidence reviewed above regarding the long standing history supporting a distinction between short-term and long-term stores, on the one hand, and recent theorizing about working memory that suggests performance on working memory tasks principally involves retrieving information from long-term memory, on the other hand.

The Present Research

A processing approach to the working memory/long-term memory distinction

The transfer-appropriate-processing theory of memory has not been previously applied to the distinction between working memory and long-term memory. Nonetheless, it would suggest that the distinction depends on the extent to which the processes involved in working memory and long-term memory tasks are similar or dissimilar. It follows that levels-of-processing effects on working memory and long-term memory tests will be similar when encoding processes and retrieval processes match, but will differ when they mismatch.

Experiment 1 was conducted to replicate our previous findings regarding levels-of-processing effects on working memory and long-term memory tasks and to extend them by employing a different paradigm. Experiment 2 was conducted to test the hypothesis that levels-of-processing effects could be observed on a working memory test if the processes involved were more similar to those involved in long-term memory tests, which was inspired by the transfer-appropriate-processing framework.

Experiment 1 extends our previous findings by addressing a potentially important methodological issue. It is possible that the lack of levels-of-processing effects on the LOP span task was due to certain aspects of the procedure. For example, the Craik and

Tulving (1975) procedure described above presented participants with an orienting question followed by a to-be-remembered word and the processing decision was made on the to-be-remembered word (see Table 1 for specific examples). In contrast, in the LOP span task, a to-be-remembered word was presented first, followed by two words that matched the preceding word in color, rhyme, or meaning, and the processing decision was made on the associated words, not the to-be-remembered word. Therefore, it is possible that the results were due to differences in the procedure. Although the research examining levels-of-processing effects on immediate and delayed memory tests reviewed above attest to the generalizability of the results, Experiment 1 used the original Craik and Tulving (1975) materials and procedure in order to replicate and extend our previous findings with respect to levels-of-processing effects on working memory tasks using established methods for investigating levels-of-processing effects.

Experiment 1: Levels-of-processing effects on immediate recall with intentional encoding and delayed recognition.

Experiment 1 used the same procedure and stimuli as Craik and Tulving's (1975) Experiment 9. This classic levels-of-processing paradigm is known to produce robust levels-of-processing effects. The only difference with the procedure used in the present experiment was that participants performed the visual, phonological, or semantic processing decisions on groups of question-word pairs in the context of a working memory span task. That is, immediate recall was required after only a few processing decisions (3 or 8), rather than after all of the words were processed. Therefore, although the levels-of-processing task followed the same procedure as Craik and Tulving (1975)

Experiment 9, active maintenance of the target words was involved because immediate recall tests were required.

The procedure was as follows: A question was presented (i.e., “Is the following word in UPPERCASE?” for visual processing, “Does the following word RHYME with X?” for phonological processing, or “Is the following word a member of the category X?” for semantic processing), and then a word was presented. Participants needed to respond to the question by pressing a key labeled ‘Y’ or ‘N’ to indicate Yes or No, respectively. Processing decisions were made for a series of either 3 or 8 question-word pairs. After the series of decisions were made, participants were asked to recall the words on which the Yes/No decisions were made. Participants were instructed beforehand that these were “to-be-remembered” words. Thus, the condition with immediate recall testing resembled a standard complex working memory span in that answering questions was the secondary processing task. Following all of the processing decisions and immediate recall tests, participants performed 10 minutes of arithmetic problems to provide a filled retention interval. Following the 10 minutes of arithmetic, participants took a delayed recognition test on the words that were to be remembered in the initial phase of the experiment.[2]

Although this procedure is known to produce a robust levels-of-processing effect on long-term memory tests, I expected it to be eliminated on working memory tests, but to appear on a delayed recognition test involving the words from the working memory

---

[2] A delayed recognition test was administered rather than a final free recall test because recalling the 99 words that were to be remembered on the initial working memory tests might have resulted in floor level performance. Thus, for current purposes, a delayed recognition test was expected to be the most sensitive measure of long-term memory.

tests. This prediction, which was inspired by the transfer-appropriate-processing framework, arises from the hypothesis that working memory and long-term memory tests involve different encoding, maintenance and/or retrieval processes (Rose et al., 2010). In addition, delayed recognition was expected to be differentially affected by the retrieval practice provided by the initial recall tests because retrieval from secondary memory benefits long-term retention and such retrieval is hypothesized to be involved in performance on working memory tasks. If the retrieval practice provided by the working memory tests involved retrieval from secondary memory to a greater extent for recall of 8-item lists than 3-item lists, then retrieval practice should benefit long-term retention of items recalled from 8-item lists more so than items recalled from 3-item lists.

Methods

Participants and Design. In Experiment 1, twenty-four Washington University undergraduate students participated in exchange for course credit. All participants were native English speakers. The design was a 3 (Level of Processing: Visual, Phonological, Semantic) x 2 (List Length: 3- or 8-Items) x 2 (Time of Test: Immediate Recall, Delayed Recognition) within-subjects design. All variables were manipulated within-subjects. The main dependent variable was the proportion of words that were correctly recalled on the immediate recall tests and recognized as old on the delayed recognition test.

Levels-of-processing span task. Craik and Tulving's (1975) levels-of-processing paradigm was turned into a working memory span task. In this working memory task, participants were presented with a series of orienting questions that were each paired with a to-be-remembered word. The orienting question required that a decision be made about a subsequently presented word. The questions were: “Is the following word presented in UPPERCASE letters?”; “Does the following word rhyme with x?”, where x was a word that either did, or did not, rhyme with the to-be-remembered “target” word; or “Is the following word a member of the category x?”, where x was a category label (e.g., “Is the following word a type of fish?”). There were three trials each of 3-item and 8-item lists for each condition (uppercase, rhyme, category).

Procedure. Participants were tested individually at a desktop PC. The target words and orienting questions were presented visually. On each trial, a fixation cross appeared on the monitor where each target word was presented. The participant began each trial by pressing the space bar when ready, after which an orienting question was displayed for 1750 ms. After a 250 ms blank screen, a to-be-remembered target word was presented. The participant was instructed to say the word aloud, remember the word for recall at the end of the trial, and press a button labeled “Yes” or “No” in response to the orienting question. The target word remained on the screen until the participant made a response. Prior to testing, the participant was instructed to make each decision as quickly as possible without sacrificing accuracy.

After the processing decision was made, the screen was blank for 750 ms before the next target word appeared. At the end of the trial, a green box and a tone cued the participant to recall the target words aloud in the order presented. Participants were told that if they were unable to recall all of the target words, they were to recall as many as possible in the order presented. Before starting the test trials, participants performed four practice trials of 2, 3, 4, and 5 sets of target and orienting questions in order to familiarize them with the procedure. Recall responses were recorded by electronic voice recorders for later scoring. For the immediate recall test trials, participants performed three trials each of 3- and 8-item lists of target words for each level-of-processing condition. Trials for the three processing conditions (uppercase, rhyme, category) were mixed in a predetermined random order such that successive trials were not of the same condition. Prior to starting each trial, the participant was told the condition for which to base their decision. After completing all of the immediate recall tests, participants performed mental arithmetic for 10 minutes followed by a surprise recognition test.

For the recognition test, the 99 target words that were presented in the levels-of-processing span task and 99 new lure words that had never appeared in the experiment were presented individually on the computer monitor. None of the words from the orienting questions of the levels-of-processing span task were included in the recognition task, and participants were informed of this fact. Lures were matched to the target words based on length and word frequency. For each word, participants were instructed to indicate whether that word was 'old,' meaning it was presented in one of the three processing conditions, or 'new', meaning the word was never presented in the experiment. Each old word was one of the target words that was to be read aloud during the initial processing phase and was to be remembered on the immediate recall tests.

Stimuli. The stimuli used in both experiments are presented in Appendix A. The target words and orienting questions were taken from the stimuli used in Craik and Tulving (1975) Experiment 9. The present study included the same 60 orienting questions and to-be-remembered target words that were used in Experiment 9 of Craik and Tulving (1975) as well as 39 additional questions and target words. The critical difference between their procedure and that of Experiment 1 is that immediate recall was required after either 3 or 8 processing decisions.

Results

The proportion of words correctly recalled on the immediate recall tests of the levels-of-processing span task are presented in the upper half of Table 2. These data were submitted to a 3 (level of processing: visual, phonological, semantic) x 2 (list length: 3-items, 8-items) repeated measures ANOVA. The effect of level-of-processing was significant, $F(2, 46) = 20.9$, $p < .001$; however, the pattern was not as predicted by the levels-of-processing framework: overall, the shallowest (visual) processing condition ($M = .77$) was significantly better than both the phonological ($M = .68$), $F = 33.7$, $p < .001$, and the semantic ($M = .75$) processing conditions, $F = 4.2$, $p = .05$.[3] As expected, there was a main effect of list length such that a greater proportion of words were recalled from 3-item lists than 8-item lists, $F(1, 23) = 753.9$, $p < .001$. However, list length did not interact with level of processing, $F(2, 46) = 2.2$, $p = .12$. Although immediate recall of 3-item lists was at ceiling, recall of 8-item lists was within an adequate range for detecting an effect of levels of processing, yet no such effect was observed.

[3] These two-way comparisons were conducted using follow up ANOVAs, collapsing across list length.

Table 2.

Mean (SEM) Proportion of Items Correctly Recalled on the Levels-of-Processing Span Task (Intentional Encoding) and Correctly Recognized as Old on the Delayed Recognition Test for Items Initially from 3- or 8-Item Lists.

	Level of Processing		
	Visual	Phonological	Semantic
Immediate Recall			
3-Items	.99 (.01)	.92 (.02)	.98 (.01)
8-Items	.56 (.02)	.44 (.03)	.51 (.02)
Delayed Recognition			
3-Items	.61 (.05)	.66 (.04)	.73 (.04)
8-Items	.69 (.04)	.69 (.03)	.82 (.03)

Note. The false alarm rate in delayed recognition was .19 (.02).

Regarding the delayed recognition test, the proportion of words that were initially processed in 3- or 8-item lists of the levels-of-processing span task that were correctly recognized as old are presented in the bottom half of Table 2. These data were submitted to a 3 (level of processing: visual, phonological, semantic) x 2 (list length: 3-items, 8-items) repeated measures ANOVA. The effect of level of processing was significant, $F(2, 46) = 11.8$, $p < .001$, because semantically processed words were recognized better than phonologically or visually processed words. There was also a main effect of list length, $F(1, 23) = 10.3$, $p < .01$, because words initially to be remembered in 8-item lists were better recognized than were words initially to be recalled from 3-item lists. Levels of processing and list length did not interact, $F(2, 46) = 0.8$, $p = .47$.[4]

The data depicted in Figure 11 illustrate the dissociation between levels-of-processing effects on the immediate and delayed memory tests.[5] The comparison between the shallowest (visual) and deepest (semantic) processing conditions is of particular interest. Immediate recall did not differ for visually or semantically processed items. Although this may have been partly because performance was at ceiling in the case of the 3-item lists, immediate recall also did not differ between the shallowest and deepest conditions for items from the 8-item lists (see Table 2). In contrast, the delayed recognition data showed a 13% advantage of semantic processing over visual processing.

---

[4] The omnibus ANOVA was conducted in order to justify separate analysis of the immediate recall and delayed recognition data. The omnibus ANOVA with Test (Immediate Recall vs. Delayed Recognition) x LOP (visual, phonological, semantic) x List Length (3-item list vs. 8-item list) resulted in a main effect of LOP, F (2, 46) = 25.56, $p < .001$, which did not interact with test, F (2, 46) = 2.44, $p = .09$, and the three way interaction was not significant, $F < 1$.

[5] As the levels of processing effect did not interact with list length, the data in Figure 12 collapse across this factor.

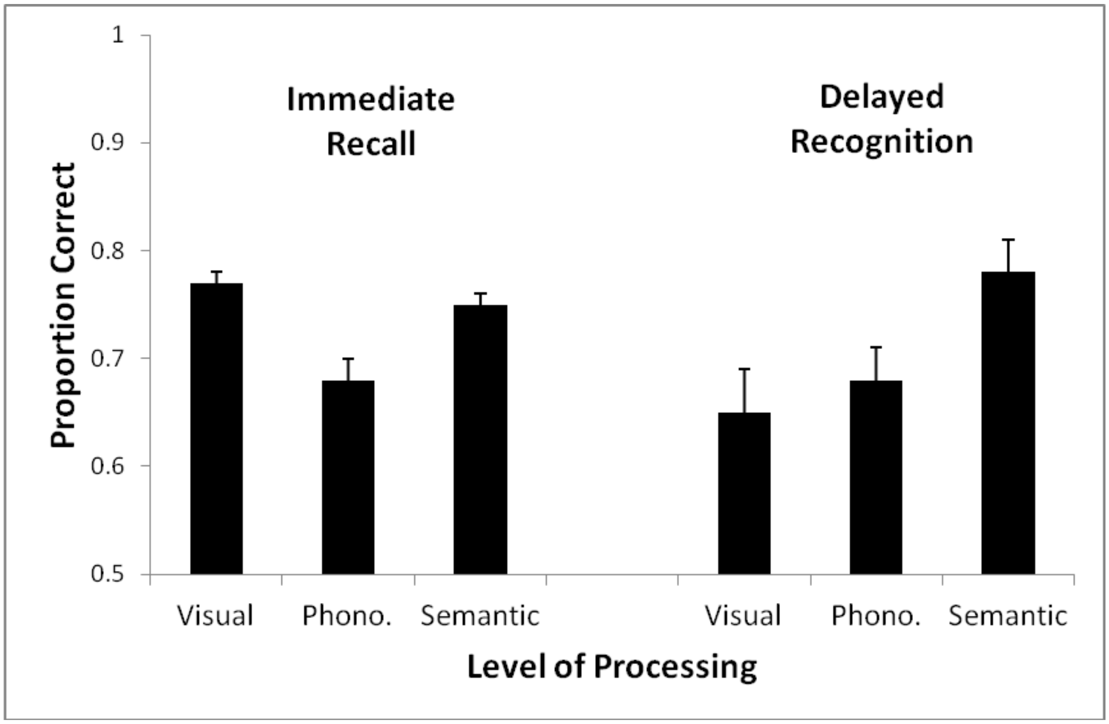


Figure 11. Mean proportion of words recalled on the immediate tests and recognized as old target words on the delayed recognition test as a function of level of processing.

Also of interest was the difference in long-term retention of items initially processed as part of 3-item versus 8-item lists (see Figure 12). Although participants recalled almost 100% of the words for 3-item lists at immediate recall, only 67 % of the words were recognized as old on the delayed recognition test. For words from 8-items lists, the opposite pattern was observed. Although participants only recalled 50% of the words at immediate recall, 73% of the words were recognized as old on the delayed recognition test.

The difference in long-term retention of items from 3- and 8- item lists points to the greater benefit of retrieval practice for items that were processed in longer, supraspan lists than for items from short, subspan lists. This finding is consistent with the hypothesis that retrieval from secondary memory was involved on the working memory span task and that retrieval of 8-item lists involved retrieval from secondary memory to a greater extent than did retrieval of 3-item lists.

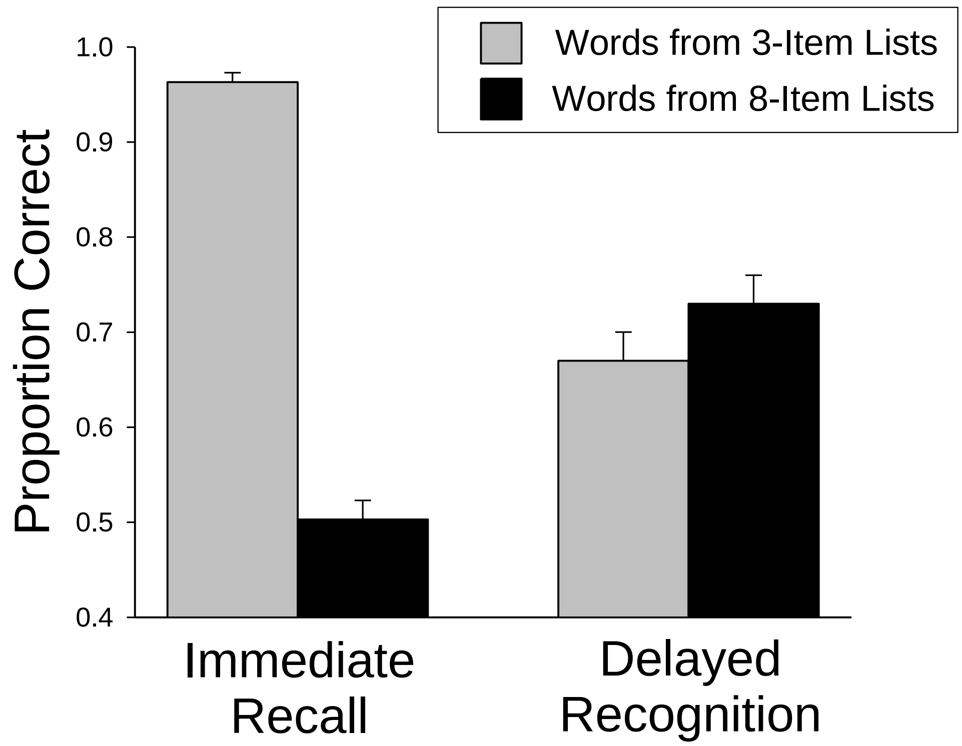


Figure 12. Mean proportion of words recalled on the immediate tests for 3- and 8-item lists and mean proportion recognized as old target words on the delayed recognition test for words initially from 3- and 8-item lists.

Discussion

The results of Experiment 1 demonstrated that, despite the use of the same levels-of-processing manipulations used by Craik and Tulving (1975) which are known to produce a very robust effect of deeper levels of processing, the effect was erased by testing memory immediately, after a few decisions, as opposed to after all of the decisions on a delayed test.

The lack of levels-of-processing effects on immediate retrieval found in Experiment 1 shows that the findings of Rose et al. (2010) were not simply due to the specific stimuli and procedures that were used in that study's methodologies. In the original version of the LOP span task used by Rose et al., the processing decision was made on the associated words, not the to-be-remembered word, whereas the current study used Craik and Tulving's original (1975) procedure which used the reverse order. Therefore, the results of Experiment 1 showed that the lack of levels-of-processing effects on the LOP span task is a robust finding, and was not simply due to a difference in the order of the levels-of-processing procedure. Additionally, the findings of Mazuryk and Lockhart (1974), Mazuryk (1974), Jacoby (1974), and Craik (1973) all point to the generalizability of dissociations between working memory and long-term memory in levels-of-processing effects.

The remaining question then is why immediate retrieval is insensitive to the effects of levels of processing? This is an especially intriguing question to address in the context of working memory research given recent evidence that suggests retrieval from secondary memory is involved in performance of working memory tasks. If this is true, then how do the processes involved in working memory and long-term memory tasks differ such that retrieval from secondary memory does not benefit from deeper levels of processing in one situation (working memory retrieval) but does in another situation (long-term memory retrieval)?

One way that working memory and long-term memory tasks may differ is in terms of the role of maintenance rehearsal processes. The performance of working memory tasks requires that information be actively maintained to prepare for an upcoming memory test. Switching attention back and forth between actively maintaining a series of words and performing other secondary processing operations is a highly demanding dual task situation. Participants may try to maintain the words by rehearsing them between performance of secondary operations and presentation of subsequent stimuli (e.g., D. P. McCabe, 2008). Active maintenance may serve to recirculate to-be-remembered words in primary memory so that, at the time of retrieval, the words are highly accessible. The active maintenance processes involved in such rehearsal may be a critical difference between immediate retrieval, as in working memory tasks, and delayed retrieval, as in long-term memory tasks.

The performance of working memory tasks involves either reporting items directly from primary memory or retrieving items from secondary memory that are highly activated because they were recently cycled through primary memory while the participant was attempting to maintain them. Thus, retrieving items on working memory tests may not benefit from deeper processing that was done at encoding; maintenance rehearsal may be sufficient. In such a situation, levels-of-processing effects are not to be expected.

The involvement of maintenance rehearsal processes may render the type of processes involved in immediate retrieval under intentional encoding conditions (e.g., standard working memory tasks) different from the processes involved in typical long-term memory tasks, even if both involve retrieval from secondary memory. If, however, retrieval was unexpected, participants would not actively maintain the words in preparation for an upcoming memory test. Rather, under incidental encoding conditions, participants would be exposed to a set of stimuli (e.g., words) and process them according to the experimenter's instructions (e.g., the level-of-processing condition). As participants process more and more information, previous items would be displaced from primary memory. Then, if an immediate recall test was administered after the items had already been processed, recalling the items on a surprise test would require retrieving them from secondary memory, but, critically, the participant would not have been attempting to actively maintain the words in mind all the while. Therefore, the type of retrieval is similar to that of the surprise long-term memory test administered after the filled retention interval, even though recall is immediate.

The comparison of levels-of-processing effects on immediate recall tests when the test is or is not expected produces an interesting prediction regarding the distinction between working memory and long-term memory. If one of the critical differences between working memory and long-term memory is the involvement of maintenance rehearsal processes that are used to keep to-be-remembered information accessible for an upcoming test, then attempting a surprise immediate recall test should show a levels-of-processing effect. Marsh, Sebrechts, Hicks and Landau (1997) reported some findings in support of this hypothesis.

Marsh et al. (1997) examined levels-of-processing effects on an adapted version of the Brown-Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959). On each trial of this task participants were presented with three words and there were three types of trials. On most trials they were to maintain the words during an unfilled retention interval and then recall the words when presented with a recall prompt. A minority of trials included a distractor-filled retention interval in which participants were required to count backwards from a random number by threes during the retention interval. On these distractor-filled trials, participants were led to believe that they would not have to recall the words. However, on a few of these distractor-filled trials a surprise recall test was administered. Because these "*critical*" trials were so infrequent (5% of the total number of trials), participants were not likely to have been expecting a recall test. When the immediate recall tests were expected, there were no differences between semantic and phonological (acoustic) levels of processing, just as in Rose et al. (2010). However, when immediate recall was unexpected (i.e., incidental encoding), semantic processing produced a significant benefit (see Figure 13).

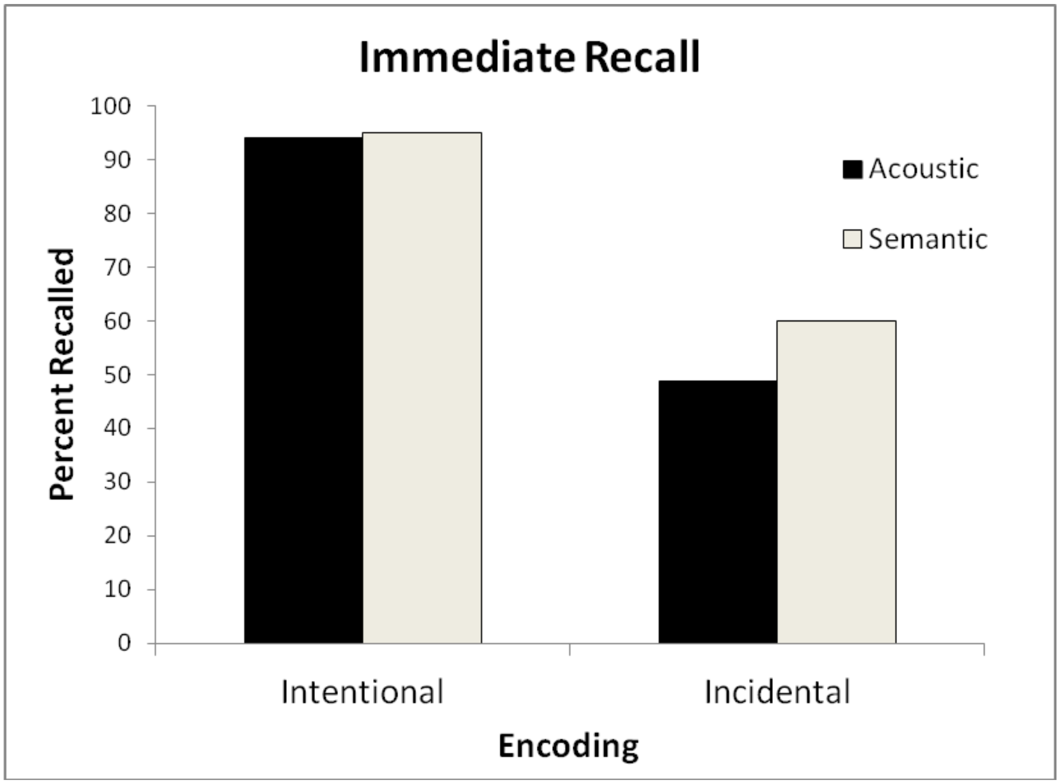


Figure 13. Immediate recall as a function of level of processing when immediate recall tests were expected (intentional encoding) and unexpected (incidental encoding). Data are estimated from Marsh, Sebrechts, Hicks and Landau (1997) Figure 1B.

The results of Marsh et al (1997) support the hypothesis that a surprise immediate recall test more closely matches the type of retrieval from secondary memory involved in delayed, long-term memory tests. Thus, the distinction between immediate retrieval in the context of a short-term or working memory task and delayed retrieval in the context of a long-term memory task may depend on the degree to which the processes involved match (or mismatch), consistent with the transfer-appropriate-processing framework.

The present study involved a second experiment in order to provide a further test of the transfer-appropriate-processing account of dissociations between levels-of-processing effects on immediate and delayed memory. Based on the predictions of the transfer-appropriate-processing account and the results of Marsh et al. (1997), Experiment 2 examined levels-of-processing effects on working memory and long-term memory for 3- and 8-item lists on surprise tests following incidental encoding.

Experiment 2: Levels-of-processing effects on immediate recall and delayed recognition when testing is unexpected.

The same general procedure used in Experiment 1 was used in Experiment 2, except that participants were not expecting the immediate recall tests. The participants were instructed to make each processing decision as quickly and accurately as they could because we were interested in the reaction time of different decisions. However, a surprise immediate recall test was administered following the final series of processing decisions. For the surprise immediate recall test, after processing the final series of words, participants were asked to try to recall as many of the words on that series as possible. Level of processing and list-length for the immediate recall test were between subjects factors: One-third of the participants were required to recall visually processed items (which was 3-items long for half of the participants and 8-items long for the other half); another third recalled phonologically processed items (either 3 or 8) and the final third recalled semantically processed items (either 3 or 8). List-length and the level of processing for the surprise immediate recall tests were manipulated between subjects so that participants were not expecting to have to recall the items on a forthcoming memory test. Therefore, the conditions involved incidental encoding of words which was then followed by a surprise immediate recall test. Because Experiment 2 assessed immediate recall of the words when testing was not expected, it was unlikely that participants would be actively maintaining the words. Therefore, I predicted that a levels-of-processing effect would appear on the immediate recall test, especially for the 8-item (supraspan) lists. If 8-item lists emphasize retrieval from secondary memory to a greater extent than 3-item lists, as was shown previously (Rose et al., 2010), then immediate recall following incidental encoding should show a larger levels-of-processing effect for 8-item lists than for 3-item lists.

Following the processing decisions and the surprise immediate recall test, just as in Experiment 1, participants performed a distractor task during a filled retention interval (10 minutes of mental arithmetic) and a delayed recognition test on the words processed in the initial phase of the experiment. Regarding delayed recognition, Experiment 2 provides a baseline, control condition with which to compare subsequent recognition of words on the levels-of-processing span task when retrieval practice for the immediate tests was or was not involved. Delayed recognition was expected to demonstrate the standard levels-of-processing effect. In addition, a testing effect was expected. That is, delayed recognition was expected to be better overall for the group that performed the immediate recall tests than the group that performed the same processing decisions, but in the context of a reaction time test.

Methods

Participants and Design. Forty eight undergraduate students participated in exchange for course credit. The design was a 3 (Level of Processing: Visual, Phonological, Semantic) x 2 (List Length: 3-Items, 8-Items) x 2 (Time of Test: Immediate Recall, Delayed Recognition) mixed design. The level-of-processing and list-length variables were manipulated between-subjects for the immediate recall test. All subjects took a final delayed recognition test on the words processed in the initial phase, making levels of processing and list length within subject variables. The main dependent variable was the proportion of words that were correctly recalled on the immediate recall tests and recognized as old on the delayed recognition test.

Procedure. The procedure was similar to that of Experiment 1 except, rather than performing the processing decisions in the context of a working memory task, participants made the same processing decisions but under the guise of a reaction time experiment. After the last trial, participants received a surprise recall test for the target words on that trial. Because this immediate recall test was unexpected, participants were probably not trying to remember the words.

Participants were instructed to make each processing decision as fast and as accurately as possible. Following a set of 3 or 8 of these decisions, a green box (which served as the recall cue for the levels-of-processing span task in Experiment 1) appeared.

Participants were instructed to pause until the next trial began. The duration of the pause was set to the mean duration that participants took to recall 3- or 8-item lists for the levels-of-processing span task in Experiment 1 (approximately 3.5 s and 10.5 s for 3- and 8-item lists, respectively). On the last trial, when the green box was displayed, an additional set of instructions appeared on the screen which read "Please repeat the words you said aloud on this trial. Try to remember as many as you can." After the surprise immediate recall test, participants performed mental arithmetic for 10 minutes followed by the recognition test, as in Experiment 1.

Results

The mean proportions of words recalled on the levels-of-processing span task on the surprise recall test are presented in the upper half of Table 3. A levels-of-processing effect was obtained on the immediate recall tests, but only for the longer (8-item) list length. That is, the deepest level of processing was best for the supraspan lists. Somewhat surprisingly, the shallowest level of processing was best for 3-item lists. These observations were statistically confirmed by an ANOVA with level of processing (visual, phonological, semantic) and list length (3-items, 8-items) as between-subjects factors. The effect of level of processing was significant, $F(2, 42) = 4.7$, $p < .05$. As expected, there was also a main effect of list length such that a greater proportion of words were recalled from 3-item lists than 8-item lists, $F(1, 42) = 35.5$, $p < .001$. In addition, list length interacted with level of processing, $F(2, 42) = 14.8$, $p < .001$, due to the fact that the deepest level of processing benefited recall of items from the 8-item lists, *F*(2, 42) = 10.7, *p* <.01, whereas the shallowest level of processing benefited recall of items from the 3-item lists, *F*(2, 42) = 8.8, *p* <.001.

The mean proportions of words from the initial levels-of-processing tasks that were later correctly recognized as old are presented in the lower half of Table 3. As can be seen, there was a clear levels-of-processing effect in that recognition was best for semantically processed items, intermediate for phonologically processed items, and worst for items that were visually processed. These observations were statistically confirmed with a repeated measures ANOVA with level of processing (visual, phonological, semantic) and list length (3-items, 8-items) as within subjects factors. The effect of level of processing was highly significant, $F(2, 94) = 100.9$, $p < .001$. This effect did not interact with list length, $F(2, 94) = 2.1$, $p = .12$, nor was there an effect of list length, $F(1, 47) = 0.9$, $p = .34$.[6]

[6] I recently conducted another study for different purposes, but which can attest to the reliability of these results. The study closely replicated the procedures of the present study, except that it was a between subjects design and a free recall test was administered to assess long-term memory for the words from the LOP span task rather than a delayed recognition test. The data are presented in Appendix C and D. These data replicate the lack of a levels-of-processing effect on the LOP span task when immediate recall was expected (Appendix C) and the appearance of such an effect when immediate recall tests were unexpected (Appendix D). In addition, subsequent delayed recall of words from the initial LOP task demonstrated LOP effects similar to the delayed recognition tests of the present study.

Table 3.

Mean (SEM) Proportion of Items Correctly Recalled on the Levels-of-Processing Span Task Following Incidental Encoding and Correctly Recognized as Old on the Delayed Recognition Test for Items Initially from 3- or 8-Item Lists.

	Level of Processing		
	Visual	Phonological	Semantic
Immediate Recall			
3-Items	.88 (.06)	.50 (.06)	.54 (.06)
8-Items	.20 (.06)	.25 (.06)	.55 (.06)
Delayed Recognition			
3-Items	.40 (.03)	.55 (.03)	.72 (.03)
8-Items	.46 (.02)	.54 (.02)	.71 (.02)

Note. The false alarm rate in delayed recognition was .14 (.01).

Discussion

As predicted by the transfer-appropriate-processing framework, a levels-of-processing effect was obtained on immediate recall, but only for 8-item lists. This finding is particularly interesting when considered alongside the results of Experiment 1. Consider, for example, the difference in levels-of-processing effects on immediate recall between Experiment 1 and 2. These data are plotted together in Figure 14. When participants made the same processing decisions on the same words, immediate recall did not show a levels-of-processing effect in Experiment 1 but did show a levels-of-processing effect in Experiment 2, at least for the supraspan (8-item) lists.

Direct comparisons should be treated with caution due to the methodological differences between Experiments 1 and 2. However, the point is that in Experiment 1 participants knew of the upcoming immediate recall test on each trial and so they were likely trying to actively maintain (rehearse) the series of words whereas in Experiment 2, participants were not expecting to have to recall the words so they would not have been maintaining them. As a result, levels of processing did not affect immediate recall in Experiment 1, but did affect immediate recall in Experiment 2. This finding is consistent with the hypothesis that, when active maintenance processes are eliminated, immediate recall on a working memory task demonstrates a levels-of-processing effect, at least for supraspan lists, similar to long-term memory tests.

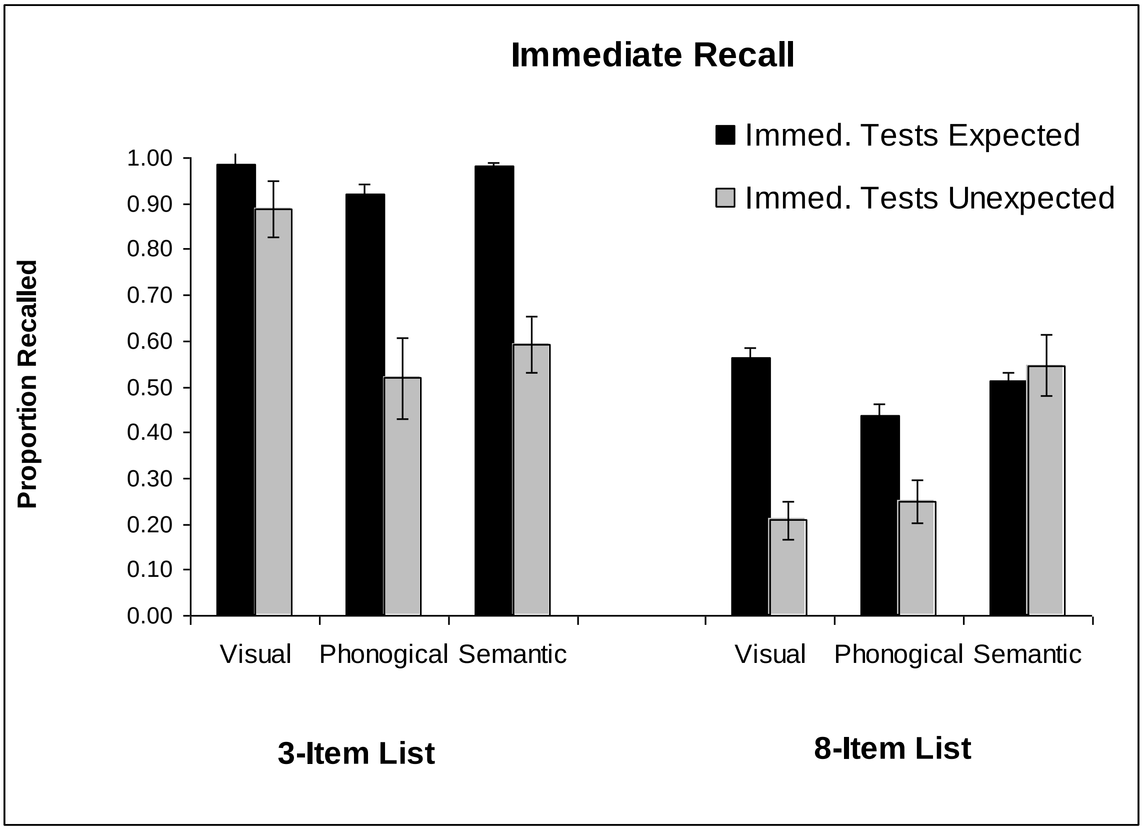


Figure 14. Mean proportion of words recalled on the immediate tests for Experiment 1 (immediate tests expected) and Experiment 2 (immediate tests unexpected) as a function of level of processing. Error bars represent the standard error of the mean.

Additionally, whether participants were expecting the immediate recall tests or not resulted in an interesting interaction between levels of processing and list length. Note that, unsurprisingly, immediate recall was better when immediate testing was expected than when it was unexpected. However, there appeared to be an interesting exception. For the deepest level of processing, recall of 8-item lists on the surprise test was as good as when immediate recall was expected (M = .55, SD = .06 vs. M = .51, SD = .02).[7]

It is also interesting to compare delayed recognition performance for Experiments 1 and 2. If trying to recall the words on the immediate recall tests (as in Experiment 1) resulted in better delayed recognition than processing the words but without immediate recall testing (as in Experiment 2), then the benefit to long-term retention would suggest that retrieval from secondary memory was involved in the immediate recall tests. This is because, as discussed above, practice retrieving items from secondary memory facilitates long-term retention whereas reporting items directly from primary memory does not (Roediger & Karpicke, 2006). Accordingly, the delayed recognition data for both Experiments are presented in Figure 15 (collapsed across list length as this factor did not interact with performance).[8]

---

[7] It should be noted that having immediate tests on every trial (as in Experiment 1) would involve much more interference than having just one immediate test (as in Experiment 2). Therefore, differences between these conditions may be also due to different amounts of proactive interference. To test this hypothesis I examined the data from Experiment 1 for both the first trial only and the last trial only, similar to what was done for Experiment 2. There was not a levels of processing effect, similar to the mean data. The data are presented in Appendix B. Therefore, the differences in immediate recall between Experiments 1 and 2 were not simply due to differences in proactive interference.

[8] For simplicity, I refer to Experiment 1 as a condition with immediate testing and Experiment 2 as a condition without immediate testing. Notably, when the delayed recognition data are analyzed excluding the three or eight words from the immediate test for Experiment 2, the pattern of results is the same as when the delayed recognition data include those three or eight words.

As can be seen, delayed recognition was better for Experiment 1 (*M* = .70) than Experiment 2 (*M* = .56). Whether or not immediate testing was required also appeared to interact with the levels-of-processing effect. Because the method for the delayed recognition test was identical for Experiments 1 and 2, the data may be directly compared. Accordingly, I conducted an ANOVA on the proportion of words recognized as old with level of processing (visual, phonological, semantic) and list length (3-items, 8-items) as within subjects factors and whether or not immediate recall was required on all trials or not as a between subjects factor. Indeed, the effect of levels of processing interacted with immediate testing, $F(2, 140) = 10.1$, $p < .001$. Follow up two-way ANOVAs showed that the interaction occurred because the benefit of immediate testing was larger for shallower levels of processing. The mean difference was .22 for the visual processing condition [$F(1, 70) = 22.2$, $p < .001$], and .13 for the phonological processing condition [$F(1, 70) = 9.9$, $p < .01$]. The difference for the semantic processing condition (.06) was not significant [$F(1, 70) = 2.7$, $p = .11$].

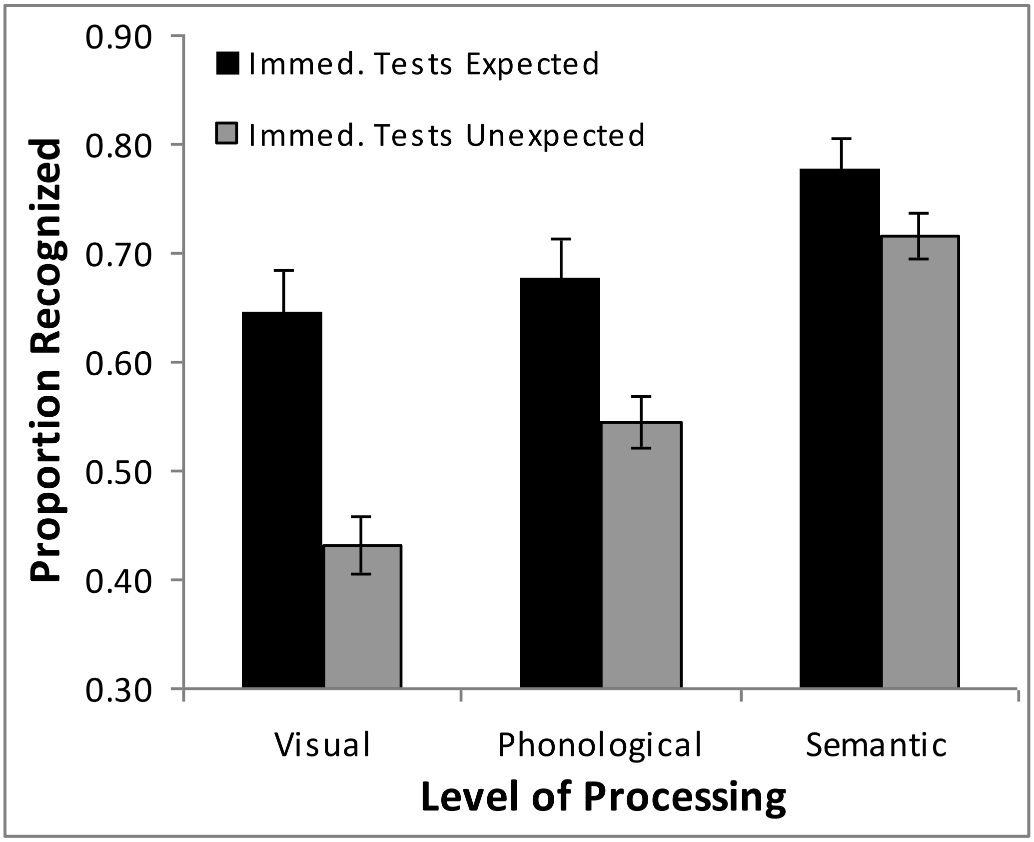


Figure 15. Mean proportion of words recognized (hits) on the subsequent delayed recognition tests when the initial immediate tests were expected (Experiment 1) and when they were unexpected (Experiment 2) as a function of level of processing, collapsed across list length.

Finally, another piece of evidence suggesting retrieval from secondary memory was involved in the LOP span task was the difference between long-term retention of words from 3- and 8-item lists which interacted with whether immediate testing was involved or not. Because participants in Experiment 2 did not have immediate recall tests, there should be no difference in long-term retention of items that were processed in 3- or 8-item lists, whereas in Experiment 1, which did have immediate recall tests, more words from the 8-item lists were recognized as old than were words from 3-item lists (.73 vs. .67, respectively). When the delayed recognition data from both Experiments were analyzed together, there was an interaction between list length and whether immediate testing was required or not, reflecting the fact that in Experiment 2, there was no difference between the 3- and 8-item lists in the proportion of items recognized (.57 vs. .56, respectively), $F(1, 70) = 4.6$, $p < .05$. This interaction suggests that, when immediate testing was required (as in Experiment 1), the retrieval practice provided by immediate testing was more beneficial for long-term retention of words from 8-item lists than words from 3-item lists. This result is what would be expected if recall of 8-item lists involved retrieval from secondary memory to a greater extent than recall of 3-item lists.

General Discussion

The goals of this study were to examine similarities and differences in levels-of-processing effects on working memory and long-term memory for the purpose of exploring the role of secondary memory in performance on working memory span tasks. In the following sections, I present an overview of the main findings and then discuss their theoretical implications.

Overview of Findings

In Experiment 1, the same paradigm that produces robust levels-of-processing effects on long-term memory tests (Craik & Tulving, 1975) was employed in the context of a working memory span task in order to test the hypothesis that working memory tasks involve retrieval from secondary memory just as in long-term memory tests. In this paradigm, participants processed to-be-remembered words based on their visual, phonological, or semantic features, and after either 3- or 8-processing decisions, they then tried to recall the words. Levels-of-processing effects were not observed on the immediate recall tests, but the effect appeared on a recognition test after a short delay involving the same words.

This finding may seem contrary to the idea that retrieval from secondary memory is involved in both working memory and long-term memory tests. However, other evidence suggests that retrieval from secondary memory was indeed involved on the initial working memory tests: Long-term retention was enhanced for words from 8-item lists relative to words from 3-item lists. This finding is consistent with the idea that retrieval practice is beneficial to the extent that the conditions require retrieval from secondary memory. That is, immediate recall of items from subspan (3-item) lists were more likely to involve direct readout from primary memory whereas immediate recall of items from supraspan (8-item) lists were more likely to involve retrieval from secondary memory, and thus provided a beneficial form of retrieval practice. Taken together, the findings of Experiment 1 demonstrate a dissociation between levels-of-processing effects on working memory tasks and long-term memory tests, despite evidence suggesting that retrieval from secondary memory was involved in performance of the working memory task.

Experiment 2 was conducted to test the hypothesis that a levels-of-processing effect would be observed on a working memory task if the processes involved in performing the task more closely matched those involved in performing a long-term memory test. Accordingly, Experiment 2 employed a similar procedure to that of Experiment 1 but with only one immediate recall test, of which participants were unaware in advance. Their task was to make the same processing decisions on the same words as fast and as accurately as possible. Then, after the last series of processing decisions had been made, participants were asked to recall the words from that series (either the last 3 or 8 words). Thus, participants were not likely to have been actively maintaining the words because they were not anticipating the need to recall them. In contrast to the immediate recall results of Experiment 1, the surprise immediate recall test of Experiment 2 demonstrated a levels-of-processing effect, but only for the 8-item lists. This result is consistent with the hypothesis that retrieval from secondary memory is involved in immediate recall of supraspan lists.

Taken together, the results of the two experiments show that the level of processing at encoding was not an important determinant of immediate recall under standard working memory conditions, but was an important determinant of immediate recall on a surprise test. The reason for this pattern of findings is likely due to the fact that on working memory tasks, participants intentionally encode to-be-remembered items and attempt to actively maintain those items in preparation for an upcoming memory test. For a surprise immediate recall test, however, participants would not attempt to actively maintain the words, making the situation similar to that in which participants get a surprise delayed recall test. This similarity in the two situations, in turn, may cause participants to use the same retrieval processes on surprise immediate tests as they use on surprise delayed tests, rather than the retrieval processes used when recalling words that are being actively maintained. Thus, these results suggest that similarities and differences between performance on working memory and long-term memory tasks depend on the extent to which the processes involved in performing the two types of tasks match or mismatch, which in turn depends on the degree of similarity between the test situations. Specifically, the results of Experiment 2 show that when intentional encoding and active maintenance processes are eliminated, immediate recall of supraspan lists demonstrates a levels-of-processing effect, similar to long-term memory tests. This novel result is consistent with a processing approach to the working memory/long-term memory distinction inspired by the transfer-appropriate-processing framework.

Implications for the theoretical distinction between working memory and long-term memory

Taken together, the present findings have implications for how memory theories should conceptualize the distinction between working memory and long-term memory. In particular, they address the anomaly created by my previous finding that immediate recall on a working memory task did not show a levels-of-processing effect (Rose et al., 2010), a result that appeared to be inconsistent with the recent hypothesis that working memory tasks involve retrieval from secondary memory (e.g., Unsworth & Engle, 2007). However, the present findings show that although there may be some differences in the processes used on working memory and long-term memory tasks, items are likely being retrieved from secondary memory on both types of tasks.

More specifically, I propose that the reason why Rose et al. (2010) observed no levels-of-processing effect with a working memory task is that, although working memory and long-term memory tests both involve retrieval from secondary memory, these tasks have very different requirements and, therefore, they recruit somewhat different cognitive processes (e.g., maintenance rehearsal in the case of working memory tests). Performing a working memory task requires maintaining a rather small set of information over the short-term. In contrast, performance on a long-term memory test involves encoding information in such a way that will facilitate its retrieval over the long-term. Of course, terms like “short-term” and “long-term” are relative. The point is that, because of their different requirements, working memory and long-term memory tests will call upon different processes. As a result, even though retrieval in both types of tests may come from the same memory system involving the same neural substrates, the two may appear to obey different principles (e.g., differential sensitivity to levels of processing).

By this account, it also follows that a levels-of-processing effect should be obtained on a working memory task if the processes involved in the initial encoding, retention, and subsequent retrieval conditions were similar to those involved in a long-term memory test. The results of Experiment 2 support this hypothesis: Immediate recall of a supraspan list of words on a surprise test did demonstrate a levels-of-processing effect. In this situation, because immediate recall was unexpected, the type of encoding and retention processes that were involved did not include processes associated with intentional encoding and active maintenance of the to-be-remembered items. Rather, the situation likely involved cue-based retrieval processes similar to those that people use on standard long-term memory tests. That immediate recall in the context of a standard working memory task did not show a levels-of-processing effect (Experiment 1), but the surprise immediate recall test did (Experiment 2) clearly supports this hypothesis.

As just noted, the surprise immediate recall tests demonstrated a levels-of-processing effect, but only for a supraspan (8-item) list. Recall of a subspan (3-item) list did not show an LOP effect. One possible source for this pattern was differences in the amount of distraction or interference produced by the secondary processing decisions between the LOP conditions. For example, the uppercase judgments of the visual processing condition presented the same question for each to-be-remembered word whereas the rhyme and category judgments presented a unique sentence for each to-be-remembered word and, half of the time, the sentence contained a rhyme or semantically associated word. Therefore, the orienting questions in the phonological and semantic conditions likely produced more distraction or interference than they did in the visual condition. Consistent with this hypothesis, the uppercase judgments were made more quickly and accurately than the rhyme or category judgments. For example, mean reaction time and percent correct was 697 ms and 99% for the visual condition and 856 ms and 94% for the semantic condition. That uppercase judgments were made more quickly and accurately than the semantic judgments suggests that the uppercase judgments were easier than the category membership judgments. Moreover, the difference in processing times resulted in a shorter amount of time between encoding and retrieval for the visual than the semantic condition. This difference in the difficulty of the secondary tasks likely resulted in differences in the amount of distraction or interference between the visual and semantic processing conditions.

Another factor that may have affected immediate recall was that the participants were required to say the target words aloud when they were presented. Auditory presentation of memory items is known to have beneficial effects to immediate recall specifically to the most recently perceived items (Conrad & Hull, 1964, 1968; Murdock & Walker, 1969). As reviewed in the introduction, auditory presentation has a strong benefit to recall of the recency (primary memory) portion of the serial position curve, but does not affect the pre-recency (secondary memory) portion (Conrad & Hull, 1964, 1968; Murdock & Walker, 1969). Perhaps saying the target words aloud resulted in a greater benefit to immediate recall of the 3-item lists than the 8-item lists.

However, visual processing only benefited immediate recall on the surprise test for 3-item lists. In contrast, semantic processing was more beneficial to immediate recall on the surprise test for the 8-item lists. The critical point regarding the different pattern of levels-of-processing effects for the 3- and 8-item lists is that the longer list length would have involved retrieval from secondary memory to a greater extent than the shorter list length. Having a shorter retention interval and/or an auditory trace of the words may have been especially beneficial for immediate recall of the 3-item lists. As reviewed in the Introduction section, very short retention intervals and auditory input are very important factors for reporting items directly from primary memory, and recalling the last three words that were perceived and spoken aloud (i.e., a 3-item list) would not have exceeded the assumed capacity of primary memory, according to the estimates of some researchers (Cowan, 1999, 2005). That deeper processing would benefit recall of items assumed to have been retrieved from secondary memory but not items assumed to have been reported directly from primary memory is consistent with previous findings. Indeed, several studies have shown a lack of benefit from semantic processing for recall of items from the recency (primary memory) portion of the serial position curve while, at the same time showing that semantic processing benefits recall of pre-recency (secondary memory) items (e.g., Seamon & Murray, 1976; Smith, Barresi, & Gross, 1971).

Seamon and Murray (1976) had participants process lists of 60 words in one of three conditions: intentional encoding, incidental encoding that involved shallow processing, or incidental encoding that involved deep processing. The intentional encoding group was instructed to remember the words for immediate recall. The incidental encoding groups were told that they were participating in an experiment on decision making. The deep (semantic) processing condition required that, for each word that was presented, the participant had to decide if the word was a general (e.g., tool) or specific (e.g., hammer) instance of a category. Subjects in the structural processing condition were given the same words but were told to focus on the position of their lips while subvocally repeating each word during presentation of the list of words, and to decide if their lips touched at the beginning (e.g., mane), the end (e.g., tomb), both the beginning and the end (e.g., bomb), or not at all (e.g., clock). The average proportion of words recalled was .29 for the intentional encoding group, .22 for the incidental encoding group with deep (semantic) processing, and .10 for the incidental encoding group with shallow (structural) processing. Of interest for present purposes was the effect of deep versus shallow processing on recall of items from the recency versus the pre-recency parts of the list. For recall of recency items, there were no differences between the deep and shallow processing groups. In contrast, for recall of pre-recency items, the deep processing group showed a slight primacy effect whereas the shallow processing group showed floor levels of recall.

The results of Seamon and Murray (1976) and those of a similar study (Smith et al, 1971) are presented in Table 4. In the study by Smith et al., participants were presented with 13 noun pairs and they had to process the words under imagery (deep processing) or rehearsal (shallow processing) instructions. Immediately following presentation of each list, the first word of one of the pairs was presented and the subject was to recall the second word of the pair. The results showed that imagery instructions benefited recall of pre-recency (secondary) memory items while rehearsal instructions benefited recall of recency (primary) memory items. This finding is similar to the interaction between levels of processing and list length in Experiment 2 which showed that deep (semantic) processing benefited recall of the 8-item list, but shallow (case) processing benefited recall of the 3-item list.

Although Smith et al. (1971) suggested that a tradeoff may occur with levels-of-processing effects between primary and secondary memory, factors similar to the ones discussed above (e.g., shorter retention intervals for shallower processing conditions, auditory vocalization of to-be-remembered words) may have also contributed to the findings of the Smith et al. study. In their study, rehearsal instructions required that participants say the word aloud whereas imagery instructions did not. The fact that recall of recency items in the rehearsal condition was better than the imagery condition could have been due to the rehearsal condition involving vocalization of the words whereas the imagery condition did not.

Table 4.

Mean proportion of words estimated to have been recalled from primary (recency) and secondary (pre-recency) memory in Seamon and Murray (1976) and Smith et al. (1971).

Level of Processing Instructions	Primary Memory (Recency)	Secondary Memory (Pre-recency)
Seamon & Murray (1976)		
Structural	.06	.05
Semantic	.06	.16
Smith et al. (1971)		
Rehearsal	.81	.45
Imagery	.47	.66

Therefore, the results of the present study showing a benefit of semantic processing to incidental recall of 8-items lists, but not 3-item lists is consistent with previous results showing that semantic processing is especially beneficial for retrieval conditions that emphasize recall from secondary (i.e., recall of pre-recency items in supraspan lists; Seamon & Murray, 1976; Smith et al., 1971).

What is more important is the difference in levels-of-processing effects on immediate recall between Experiments 1 and 2. The present study showed that there was not a levels-of-processing effect in the context of standard working memory conditions (Experiment 1). This was true even for 8-item lists that exceeded working memory capacity and so, by definition, retrieval from secondary memory was likely involved. Although recent memory theories suggest that performance on working memory tests involves retrieval from secondary memory, the pattern of levels-of-processing effects on working memory and long-term (secondary) memory tests were different. The transfer-appropriate-processing framework suggests that such differences may be due to the involvement of different encoding, maintenance, and/or retrieval processes (as opposed to the involvement of different memory systems).

The findings of Experiment 2, which are consistent with the transfer-appropriate-processing account of the working memory/long-term memory distinction, show that immediate recall of supraspan lists did show a levels-of-processing effect on surprise recall tests. This suggests that, when participants were not actively maintaining the words because they were not expecting an immediate recall test, immediate and delayed retrieval demonstrated similar effects of levels of processing, suggesting the nature of retrieval was more similar.

Therefore, the key theoretical implication of the present findings is that, in contrast to recent theories suggesting that performance on both working memory and long-term memory tasks principally rely on retrieval from secondary memory and demonstrate similar retrieval dynamics (e.g., Unsworth & Engle, 2007), performance on working memory span tasks and long-term memory tests under standard conditions demonstrate different principles (e.g., sensitivity to levels of processing). Yet, consistent with the transfer-appropriate-processing account of the working memory/long-term memory distinction, if the processes involved in performance on working memory and long-term memory tasks are made to be more similar, then the two will demonstrate similar principles.

Concluding Remarks

Models of working memory must account for the way levels-of-processing effects interact between immediate retrieval, as required by working memory span tasks, and delayed retrieval, as required by long-term memory tasks. The results of the present study may be accommodated by the transfer-appropriate-processing account of the working memory/long-term memory distinction. The transfer-appropriate-processing account should serve as a useful guide for clarifying current theorizing – and expanding future theorizing – about the nature of working memory, long-term memory, and the relation between the two.

Appendix A.

Stimuli

Orienting Question (Is the following word …)	Target Word	Correct Response
a human expression?	PLATE	No
a wild animal?	bear	Yes
a type of chicken?	TENT	No
in uppercase?	FIDDLE	Yes
in uppercase?	tongue	No
in uppercase?	PIPE	Yes
in uppercase?	child	No
in uppercase?	TRUCK	Yes
in uppercase?	bike	No
in uppercase?	CHAPEL	Yes
in uppercase?	WITCH	Yes
a rhyme of shrug?	BUG	Yes
a rhyme of screech?	pine	No
a rhyme of bin?	GRIN	Yes
a type of bird?	LARK	Yes
a type of vehicle?	queen	No
something used for sleep?	BED	Yes
a type of water sport?	EARL	No

something in a park?	bench	Yes
a part of a car?	cloud	No
a type of material?	WOOL	Yes
a type of city?	CHARM	No
in uppercase?	knife	No
in uppercase?	BREAD	Yes
in uppercase?	knee	No
a rhyme of wife?	brake	No
a rhyme of flood?	MUD	Yes
a rhyme of coach?	rock	No
a rhyme of breezy?	COAL	No
a rhyme of again?	hen	Yes
a rhyme of lush?	brush	Yes
a rhyme of type?	GLASS	No
a rhyme of feet?	moan	No
in uppercase?	LAMP	Yes
in uppercase?	boat	No
in uppercase?	GAS	Yes
a rhyme of camp?	JADE	No
a rhyme of rote?	DAISY	No
a rhyme of peak?	week	Yes
a rhyme of shield?	FIELD	Yes

a rhyme of crass?	pail	No
a rhyme of ringer?	twig	No
a rhyme of leap?	SHEEP	Yes
a rhyme of ache?	rake	Yes
a type of shoe?	boot	Yes
a type of insect?	SON	No
a type of criminal?	robber	Yes
in uppercase?	church	No
in uppercase?	STREET	Yes
in uppercase?	trout	No
in uppercase?	THROAT	Yes
in uppercase?	guest	No
in uppercase?	CLIP	Yes
in uppercase?	clove	No
in uppercase?	cheek	No
a rhyme of start?	cart	Yes
a rhyme of search?	nurse	No
a rhyme of young?	SONNET	No
a part of an animal?	claw	Yes
a type of farm animal?	FLOUR	No
a type of game?	pool	Yes
a type of tree?	CAVE	No

a type of grain?	RICE	Yes
associated with medicine?	bride	No
a type of precious stone?	DRAIN	No
a type of occupation?	miner	Yes
a rhyme of female?	SAIL	Yes
a rhyme of instead?	copper	No
a rhyme of mourn?	corn	Yes
a division of time?	TRIBE	No
something hot?	flame	Yes
something to wear?	GLOVE	Yes
a type of fruit?	cherry	Yes
a type of metal?	DANCE	No
a type of dirt?	FENCE	No
a part of a ship?	mast	Yes
a type of flower?	stairs	No
in uppercase?	beach	No
in uppercase?	POND	Yes
in uppercase?	singer	No
a rhyme of noun?	town	Yes
a rhyme of crate?	STATE	Yes
a rhyme of elite?	clerk	No
a rhyme of compel?	LANE	No

a rhyme of rope?	soap	Yes
a rhyme of trunk?	MONK	Yes
a rhyme of bout?	juice	No
a rhyme of goodwill?	HILL	Yes
a territorial unit?	honey	No
a part of a room?	FLOOR	Yes
a form of communication?	speech	Yes
in uppercase?	ROACH	Yes
in uppercase?	bell	No
in uppercase?	SLEET	Yes
in uppercase?	drill	No
in uppercase?	tire	No
in uppercase?	GRAM	Yes
in uppercase?	SACK	Yes
in uppercase?	chair	No

Appendix B.

Mean (SD) Proportion of Items Correctly Recalled on the Levels-of-Processing Span Task Following Intentioanl Encoding for the First and Last Trial of the 3- and 8-Item Lists.

	Level of Processing		
	Visual	Phonological	Semantic
First Trial			
3-Items	.99 (.07)	.97 (.09)	.97 (.09)
8-Items	.60 (.14)	.41 (.18)	.57 (.15)
Last Trial			
3-Items	.99 (.07)	.96 (.11)	.99 (.07)
8-Items	.58 (.22)	.51 (.17)	.48 (.18)

Appendix C.

Mean (SEM) Proportion of Items Recalled on the Immediate and Delayed Tests as a Function of Levels-of-Processing List Length.

	Level of Processing		
	Visual	Phonological	Semantic
Immediate Recall			
3-Items	.96 (.02)	.94 (.02)	.90 (.02)
8-Items	.69 (.03)	.49 (.03)	.46 (.03)
Delayed Recall			
3-Items	.25 (.03)	.16 (.03)	.26 (.03)
8-Items	.16 (.03)	.14 (.03)	.25 (.03)

Appendix D.

Mean (SEM) Proportion of Items Correctly Recalled on the Levels-of-Processing Span Task Following Incidental Encoding and on the Delayed Test for Items Initially from 3- or 8-Item Lists.

	Level of Processing		
	Visual	Phonological	Semantic
Immediate Recall			
3-Items	.93 (.04)	.75 (.04)	.90 (.04)
8-Items	.31 (.04)	.24 (.04)	.38 (.04)
Delayed Recall			
3-Items	.16 (.02)	.13 (.02)	.24 (.02)
8-Items	.18 (.01)	.15 (.01)	.22 (.01)

References

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation: II*. Oxford England: Academic Press.

Baddeley, A., & Warrington, E. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning & Verbal Behavior, 9,* 2, 176-189.

Baddeley, A. D. (1986). *Working memory.* New York, NY, US: Clarendon Press/Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation, 8,* (pp. 47–90). New York: Academic Press.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 36A, 2, 233-252.

Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology 10*, 12-21.

Campoy, G., & Baddeley, A. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16*, 4, 329-340.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 4, 429-432.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10,* 4, 135-136.

Cowan, N. (1999). An embedded-processes model of working memory. In: Miyake, A., Shah, P. (Eds); *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 62-101). New York, NY, US : Cambridge University Press

Cowan, N. (2005). Working memory capacity. New York, NY, US: Psychology Press.

Craik, F. I. M. (1968). Two components in free recall. *Journal of Verbal Learning & Verbal Behavior, 7,* 6, 996-1004.

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning & Verbal Behavior, 9,* 2, 143-148.

Craik, F. I. M. (1973). A "levels of analysis" view of memory. In P. Pliner, L. Krames, & T. M. Alloway (Eds.), *Communication and affect: Language and thought*. London: Academic Press.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future?. *Memory, 10,* 5, 305-318.

Craik, F. I. M., & Jacoby, L. L. (1975). A process view of short-term retention. In F. Restle, R.N. Shiffrin, H.J. Castellan, M.R. Lindman & D.B. Pisoni (Eds.), *Cognitive Theory, 1,* (pp. 171-192). Potomac, MD: Erlbaum.

Craik, F. I. M., & Levy, B. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology, 86*, 1, 77-82.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General. 104*, 3, 268-294.

Crowder, R. G. (1968). Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review, 85*, 6, 502-524.

Daneman, M., & Carpentar, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 4, 450-466.

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports, 5,* 305-312.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports, 7,* 337-344.

Ebbinghaus, H. E. (1885/1964). *Memory: A contribution to experimental psychology*. Transl. HA Ruger, CE Bussenius, 1913. New York: Dover.

Glanzer, M., & Cunitz, A. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, *5*, 4, 351-360.

Jacoby, L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning & Verbal Behavior, 13*, 5, 483-496.

James, W. (1890). *The principles of psychology*. New York: Holt, Rinehart & Winston.

Jonides, J., Lewis, R., Nee, D., Berman, M., Moore, K., & Lustig, C. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193-224.

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language. 57,* 2, 151-162.

Craik, F., & Levy, B. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*, *86,* 1, 77-82.

Madigan, S., & McCabe, L. (1971). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 101–106.

Marsh, R., Sebrechts, M., Hicks, J., & Landau, J. (1997). Processing strategies and secondary memory in very rapid forgetting. *Memory & Cognition, 25*, 2, 173-181.

Mazuryk, G. (1974). Positive recency in final free recall. *Journal of Experimental Psychology, 103*, 4, 812-814.

Mazuryk, G. F., & Lockhart, R. S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology*, *23,* 114-123.

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language, 58,* 480-494.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27,* 817–835.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation,* 46. San Diego: Academic Press.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 2, 81-97.

Milner, B. (1966). Amnesia following operation on the temporal lobes. In C. W. M. Whitty &. O. L. Zangwill (Eds.), *Amnesia*. (pp. 109–133). London: Butterworths.

Morris, C., Bransford, J., & Franks, J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*, 5, 519-533.

Murdock, B., & Walker, K. (1969). Modality effects in free recall. *Journal of Verbal Learning & Verbal Behavior, 8,* 5, 665-676.

Murdock, B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 5, 482-488.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 3, 411-421.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 61*, 12-21.

Postman L. & Phillips L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology. 17,* 132-138.

Roediger, H. L., III, & Crowder, R. G. (1976). Recall instructions and the suffix effect. *American Journal of Psychology, 89*, 115-125.

Roediger, H. L., III. & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives in Psychological Science, 1*, 3, 181-210.

Rose, N. S., Myerson, J., & Roediger, H. L., (in preparation). Individual differences in working memory, secondary memory, and fluid intelligence: Evidence from the levels-of-processing span task.

Rose, N. S., Myerson, J., Roediger, H. L., & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 2, 471-483.

Seamon, J., & Murray, P. (1976). Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory*, *2,* 6, 680-687.

Shallice, T., & Warrington, E. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology*, *22*, 2, 261-273.

Smith, E., Barresi, J., & Gross, A. (1971). Imaginal versus verbal coding and the primary-secondary memory distinction. *Journal of Verbal Learning & Verbal Behavior, 10*, 6, 597-603.

Stein, B. S. (1978). Depth of processing reexamined: The effects of the precision of encoding and test appropriateness. *Journal of Verbal Learning & Verbal Behavior, 17*, 2, 165-174.

Tulving, E., & Patterson, R. (1968). Functional units and retrieval processes in free recall. *Journal of Experimental Psychology*, *77*, 2, 239-248.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, *5*, 4, 381-391.

Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 5, 352-373.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54*, 68–80.

Unsworth, N., & Engle, R. W. (2007). The Nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review. 114*, 1, 104-132.

Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin, 81*, 10, 695-711.

Waugh, N., & Norman, D. (1965). Primary memory. *Psychological Review, 72*, 2, 89-104.